BIRTH of the COSMOS field: primordial and evolved density reconstructions during cosmic high noon

Metin Ata,[1]★† Francisco-Shu Kitaura,[2,3] Khee-Gan Lee,[1] Brian C. Lemaux [ID],[4] Daichi Kashino [ID],[5] Olga Cucciati [ID],[6] Mónica Hernández-Sánchez[2,3] and Oliver Le Fèvre[7]

[1]*Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), WPI, The University of Tokyo Institutes for Advanced Study (UTIAS), The University of Tokyo, Kashiwa, Chiba 277-8568, Japan*
[2]*Instituto de Astrofísica de Canarias (IAC), Calle Vía Lactea s/n, E-38200 La Laguna, Tenerife, Spain*
[3]*Departamento de Astrofísica, Universidad de La Laguna (ULL), E-38206 La Laguna, Tenerife, Spain*
[4]*Department of Physics, University of California, Davis, One Shields Ave., Davis, CA 95616, USA*
[5]*Department of Physics, ETH Zürich, Wolfgang-Pauli-Strasse 27, CH-8093 Zürich, Switzerland*
[6]*INAF - Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Gobetti 93/3, I-40129 Bologna, Italy*
[7]*CNRS, LAM (Laboratoire d'Astrophysique de Marseille), Aix Marseille Université, UMR 7326, F-13388 Marseille, France*

ABSTRACT

This work presents the first comprehensive study of structure formation at the peak epoch of cosmic star formation over $1.4 \leq z \leq 3.6$ in the Cosmic Evolution Survey (COSMOS) field, including the most massive high-redshift galaxy proto-clusters at that era. We apply the extended COSMIC BIRTH algorithm to account for a multitracer and multisurvey Bayesian analysis at Lagrangian initial cosmic times. Combining the data of five different spectroscopic redshift surveys (zCOSMOS-deep, VUDS, MOSDEF, ZFIRE, and FMOS–COSMOS), we show that the corresponding unbiased primordial density fields can be inferred, if a proper survey completeness computation from the parent photometric catalogues, and a precise treatment of the non-linear and non-local evolution on the light-cone is taken into account, including (i) gravitational matter displacements, (ii) peculiar velocities, and (iii) galaxy bias. The reconstructions reveal a holistic view on the known proto-clusters in the COSMOS field and the growth of the cosmic web towards lower redshifts. The inferred distant dark matter density fields concurrently with other probes like tomographic reconstructions of the intergalactic medium will explore the interplay of gas and dark matter and are ideally suited to study structure formation at high redshifts in the light of upcoming deep surveys.

Key words: galaxies: high-redshift – large-scale structure of Universe – cosmology: theory.

1 INTRODUCTION

Our current standard cosmological model predicts a hierarchical clustering of subsequently merging small structures to greater ones (see Fry & Peebles 1978; White & Rees 1978) up to the formation of the largest galaxy superclusters observed in our present local Universe (e.g. Tully et al. 2014, with a total mass of $\sim 10^{17}$ M$_\odot$). The formation history and the exploration of the underlying physical phenomena of galaxy clustering over cosmic history remains an important question (see the pioneering work of Kauffmann et al. 1999, and references therein).

Furthermore, progenitors of galaxy clusters and superclusters and their haloes are key probes to understand early structure formation (Gao et al. 2005; Cohn & White 2008) and can be used to constrain a particular dark matter model (Bode, Ostriker & Turok 2001). Also, the analysis of galaxy clusters allow us to study galaxy formation, test cosmological parameters (Allen, Evrard & Mantz 2011) and constrain non-standard cosmological models (Kravtsov & Borgani 2012; Costanzi Alunno Cerbolini et al. 2013).

In particular, the range of $2 \lesssim z \lesssim 3$ marks the peak epoch of star formation in the Universe, frequently referred to as 'Cosmic High Noon' (Somerville & Davé 2015). The processes driving star formation of galaxies and clusters have been studied in previous works, showing a nontrivial relation of star formation and the galaxies' environmental density (Cooper et al. 2008; Koyama et al. 2013; Kawinwanichakij et al. 2017; Ji et al. 2018; Muldrew, Hatch & Cooke 2018). Moreover, the quenching of star formation within massive galaxy clusters has been found (Cooper et al. 2008), suggesting on average a passive evolution of galaxy clusters during the last ≈10 Gyr and studying an environmental dependence (see Lemaux et al. 2012; Belfiore et al. 2017; Lemaux et al. 2019; Tomczak et al. 2019). Therefore, an accurate description of the dark matter density distribution at these redshifts will potentially help to understand the interconnection of star formation (see e.g. Behroozi, Wechsler & Conroy 2013; Madau & Dickinson 2014; Shi et al. 2019) and the location within the cosmic web (e.g. Bond, Kofman & Pogosyan 1996). Detailed analyses of high-redshift structures have been accessible via numerical simulations (e.g. Watson et al. 2013; Henriques et al. 2014; Chiang et al. 2017), studying their formation history and the importance of the environment.

While wide field galaxy surveys like SDSS-III Baryon Oscillation Spectroscopic Survey Alam et al. 2017) and 2MASS (Skrutskie et al.

★ E-mail: metin.ata@ipmu.jp
† Kavli IPMU Fellow.

2006) have played an important role in spatially mapping the large-scale structure of galaxies and quasars (see e.g. Ata et al. 2018), these have focused on scales of the baryon acoustic oscillations (Eisenstein et al. 2005; Percival et al. 2007). Some endeavours have been done to map the structure formation down to megaparcec scales over to smaller footprints in so-called pencil beams,[1] such as ALMA deep fields (Casey et al. 2018) reaching out to $z \sim 1.5–2.5$. However, no spectroscopic surveys have abundantly mapped the large-scale density distribution down to megaparsec scales beyond $z > 1.5$, although multiple projects aim to push beyond this boundary in the near future, such as DESI (Levi et al. 2013), *EUCLID* (Amendola et al. 2018), PFS (Takada et al. 2014), 4MOST (de Jong et al. 2012), and MOONS (Cirasuolo et al. 2014). For this reason, we are still missing observational validation of quasi-linear structure formation at about $z \gtrsim 1.5$, before non-linearities start dominating the emerging cosmic web.

Over nearly two decades, the Cosmic Evolution Survey (COSMOS; see latest version Capak et al. 2007; Scoville et al. 2007; Laigle et al. 2016) has been a major ongoing effort to photometrically observe galaxies across a sufficiently wide footprint to resolve transverse large-scale structure, while simultaneously having sufficient depth to probe the entire span of cosmic history from the Local Universe right into (and eventually beyond) the Epoch of Reionization (e.g. Scoville et al. 2013). Observations within COSMOS have identified large aggregations of galaxies at high redshifts (Yuan et al. 2014), i.e. galaxy proto-clusters, that have been extensively studied over a large redshift range at Cosmic Noon (Diener et al. 2013, 2015; Casey et al. 2015; Chiang et al. 2015; Lee et al. 2016; Wang et al. 2016; Cucciati et al. 2018; Lemaux et al. 2018; Darvish et al. 2020). These overdense structures have been proposed to be possible progenitors of the Coma-like galaxy clusters, potentially assuming total masses of $\sim (1–2) \times 10^{15}\,M_\odot$ at $z = 0$ (Lee et al. 2016; Cucciati et al. 2018; Lemaux et al. 2018; Darvish et al. 2020). The COSMOS field is therefore ideally suited to study early structure formation and the evolution of galaxy proto-clusters into the mature structures we observe at the current epoch.

However, some of the aforementioned pioneering studies have typically been based on individual spectroscopic surveys in the COSMOS field and also not taking selection criteria into account. Therefore, a consistent analysis of the confirmed structures combining the multiple deep galaxy spectroscopic surveys over a sufficiently large redshift range is still missing. This has led to a heterogeneous view of structures that may or may not co-evolve, e.g. the reported overdensities at $z \approx 2.4–2.5$ that are within $\sim 100\,h^{-1}$ Mpc of each other (Casey et al. 2015; Chiang et al. 2015; Diener et al. 2015; Wang et al. 2016). Moreover, analysis of the overdensities have typically been carried out by comparing with analogous structures in *N*-body simulations rather than direct analysis of the observed structures, which leads to greater uncertainties due to the diversity of structures with similar aggregate properties in the simulations (e.g. aperture mass, velocity dispersion, etc). This situation compounds an attempt to compare the results of these findings.

Our goal in this work is to recover the dark matter density distribution in the COSMOS field during Cosmic Noon ($1.4 \leq z \leq 3.6$), jointly constrained from different spectroscopic galaxy surveys, revealing a consistent reconstruction of all structures within

the field. We infer the Gaussian density field at redshift $z = 100$ and the density field corresponding to the observed redshift over the range of $1.4 \leq z \leq 3.6$, which we refer to as 'initial' and 'final' conditions, respectively. Over the last two decades, several density reconstruction methods have been proposed in literature, starting from iterative methods (Zaroubi et al. 1995; Kitaura et al. 2009; Wang et al. 2009), over non-linear ones using, e.g. a lognormal prior and a Poisson likelihood (Kitaura, Jasche & Metcalf 2010), to more realistic structure formation models, such as Lagrangian perturbation theory (LPT; Jasche & Wandelt 2013; Kitaura 2013; Wang et al. 2013; Schmittfull, Baldauf & Zaldarriaga 2017; Hada & Eisenstein 2018; Kitaura et al. 2019; Lavaux, Jasche & Leclercq 2019; Patrick Bos, Kitaura & van de Weygaert 2019) or Particle-Mesh based codes (Wang et al. 2014; Jasche & Lavaux 2019). Recent reconstruction approaches also aim to infer initial and final conditions from the observations of absorption lines from the intergalactic medium (IGM; see e.g. Kitaura, Gallerani & Ferrara 2012a; Horowitz et al. 2019; Porqueres et al. 2019).

In this work, we use five spectroscopic galaxy surveys in the COSMOS field: (i) zCOSMOS-deep (Lilly et al. 2009), (ii) VI-MOS Ultra Deep Survey (VUDS, Le Fèvre et al. 2015), (iii) MOSFIRE Deep Evolution Field (MOSDEF, Kriek et al. 2015), (iv) KECK/MOSFIRE Spectroscopic Survey of Galaxies in Rich Environments (ZFIRE, Nanayakkara et al. 2016), and the (v) FMOS–COSMOS survey (Silverman et al. 2015; Kashino et al. 2019) to jointly reconstruct the initial dark matter density field with the COSMIC BIRTH method (Kitaura et al. 2019), which is ideally suited for this purpose as it deals for the first time with light-cone evolution effects beyond the Zel'dovich approximation. Considering these surveys, we first develop a formalism to combine them within our Bayesian reconstruction framework. Secondly, we calculate the selection functions of each survey to estimate the completeness of the observations as a function of the angular and radial dimensions. Thirdly, we study the galaxy bias beyond passive evolution as a function of redshift.

The inferred density fields are ideally suited to compare with the IGM absorption maps, from the $z \sim 2–2.5$ CLAMATO (COSMOS Lyα Mapping And Tomographic Observations) survey (Lee et al. 2014a,b, 2016, 2018), and the $z \sim 2.2–2.8$ LATIS (Lyα Tomography IMACS Survey; Newman et al. 2020), based on the Lyα forest tomography technique (Pichon et al. 2001) and thus will provide insights into the relationship of matter clustering and the properties of the IGM. Other future applications include a direct study of galaxy properties as a function of underlying matter density and a full constrained *N*-Body simulation starting from the inferred initial conditions.

This article is structured as follows. In Section 2, we give a brief introduction to the recently developed initial density perturbations reconstructions algorithm COSMIC BIRTH. We discuss the challenges to reconstruct the density perturbations in the COSMOS field, and present the necessary extensions to our algorithm. In Section 3, we give details of the surveys used in this work, how we selected the data and how we constructed the selection functions. In Section 4, we give details of our large-scale bias computation and how we determined the bias evolution with increasing redshift. In Section 5, we describe the application of COSMIC BIRTH to the surveys in the COSMOS field, present the inferred density fields and provide additional diagnostics to validate the COSMIC BIRTH reconstruction algorithm. Finally, we give a summary in Section 6, discuss our results and forecast future efforts based on this work.

---

[1]Pencil beam surveys do not statistically resolve the transverse large-scale structures at $\sim 10\,h^{-1}$ Mpc scales.

3196 M. Ata et al.

Throughout this paper, we use a fiducial flat ΛCDM cosmology with a set of cosmological parameters $\{p_c\}$ of

$$\{p_c\} = \{\Omega_M = 0.31, \Omega_\Lambda = 0.69, \sigma_8 = 0.82, n_s = 0.96, h = 0.68\} \tag{1}$$

concertedly chosen with the parameters in Lee et al. (2018). All distances are given in comoving h^{-1} Mpc units.

2 COSMIC BIRTH-ALGORITHM

Density field reconstructions from observed galaxy distributions are an ongoing effort in cosmological science. The aim of this work is to infer the initial density fluctuations $\delta(q)$ at Lagrangian coordinated q given galaxy positions that have been observed in Eulerian redshift-space s. These two frames can be connected if the peculiar velocity $v_r(q)$ and displacement $\psi(q)$ fields are known, as

$$q = s - v_r(q) - \psi(q). \tag{2}$$

Equation (2) shows that the mapping problem is analytically ill-defined, as we need a priori knowledge of the Lagrangian coordinates q at which the displacement field and the peculiar velocity field (see Kitaura et al. 2016) are evaluated. Thus, an iterative solution is required to solve this problem, as proposed in pioneering works (Yahil et al. 1991; Monaco & Efstathiou 1999). This class of iterative mapping schemes were further developed in Kitaura & Angulo (2012), Kitaura et al. (2012b, 2019), Kitaura (2013), and Heß, Kitaura & Gottlöber (2013). One of the key ingredients to achieve higher accuracy on small scales consists of introducing a Bayesian inference framework in the initial conditions reconstruction that takes into account the likelihood of the dark matter tracers. In this way, the typical additional Gaussian smoothing can be avoided and the number counts of objects within the mesh in a given voxel resolution can be correctly treated (Kitaura & Enßlin 2008; Jasche & Kitaura 2010; Kitaura et al. 2010). This allows for high precision on a few Mpc scales (see Kitaura et al. 2012c; Nuza et al. 2014).

COSMIC BIRTH (Kitaura et al. 2019) is a Bayesian inference framework, using a nested Gibbs-sampling scheme to infer the initial density perturbations $\delta(q)$ on a regular cubical mesh grid with N_C voxels. It maps the galaxy distribution represented by its Cartesian (for a given set of cosmological parameters) Eulerian redshift-space positions $\{s^{obs}\}$ to Lagrangian real-space coordinates $\{q\}$ expressed as number counts of galaxies per voxel on a regular mesh: $N_G(q)$. For efficiency, the displacement field is computed relying on Augmented Lagrangian perturbation theory (Kitaura & Hess 2013) and so effectively solving equation (2; see Section 2.1).

In this way, we have a precise description of the action of gravity within ΛCDM on Mpc scales at redshifts larger than one (Neyrinck 2013). To compute the displacement $\psi(q)$ and velocity fields $v_r(q)$ from the initial density field $\delta(q)$ we apply Hamilton Monte Carlo (HMC) sampling (see Duane et al. 1987; Jasche & Kitaura 2010; Neal 1993, 2011) with a bias description $B(\delta(q))$ and the galaxy counts on the mesh grid N_G. This is key to perform a forward modelling, as we do not obtain the displacement field from the density field defined in Eulerian space, done by inverse approaches (see e.g. Eisenstein et al. 2007). We account for selection effects of the galaxy survey data, i.e. the survey geometry, the angular and the radial selection functions, within a response operator R (see Section 3.2 for details). The COSMIC BIRTH code uses a particularly efficient fourth-order leap-frog implementation (see Hernández-Sánchez et al. 2019) to solve the Hamiltonian equations of motion. We recap the gravity model in Section 2.1 following a description of the probabilistic

model in more detail in Section 2.2 and then expand the calculations in Section 2.4 to combine multiple surveys.

2.1 Gravity model

In this section, we recapitulate the analytical model used within the COSMIC BIRTH code to compute the gravitational evolution of the cosmic density field.

We rely on augmented Lagrangian Perturbation Theory (ALPT) to simulate structure formation (details can be found in Kitaura & Hess 2013). In this approximation, the displacement field $\psi(q)$, mapping a distribution of dark matter particles at initial Lagrangian positions q to the final Eulerian positions $x(z)$ at redshift z ($x(z) = q + \psi(q)$), is split into a long-range $\psi_L(q)$ and a short-range component $\psi_S(q)$, i.e. $\psi(q) = \psi_L(q) + \psi_S(q)$. The long-range component is computed with second-order Lagrangian Perturbation Theory (2LPT) ψ_{2LPT} (for details on 2LPT, see Bouchet et al. 1992, 1995; Catelan 1995). The resulting displacement field is convoluted with a kernel $\mathcal{K}$: $\psi_L(q) = \mathcal{K}(q, r_S) * \psi_{2LPT}(q)$, given by a Gaussian filter $\mathcal{K}(q, r_S) = \exp(-|q|^2/(2r_S^2))$, with r_S being the smoothing radius. The short-range component is modelled with the spherical collapse approximation $\psi_{SC}(q)$ (see Bernardeau 1994; Mohayaee et al. 2006; Neyrinck 2013). The resulting ALPT displacement field from combining the long and the short range components given by

$$\psi_{ALPT}(q) = \mathcal{K}(q, r_S) * \psi_{2LPT}(q) + (1 - \mathcal{K}(q, r_S)) * \psi_{SC}(q) \tag{3}$$

is used to move a set of homogenously distributed particles from Lagrangian initial conditions to the Eulerian final ones. We then grid the particles following a clouds-in-cell scheme and phase space mapping (Abel, Hahn & Kaehler 2012; Hahn, Abel & Kaehler 2013) to produce a smooth density field $\delta(r)$. Some improvements can be obtained preventing voids within larger collapsing regions, which essentially extends these regions towards moderate underdensities (see MUSCLE method in Neyrinck 2016). This approach requires about eight additional convolutions being about twice as expensive, as the approach used here. Moreover, we have checked that the improvement provided by including MUSCLE is not perceptible when using grids with cell resolutions of the order $\sim h^{-1}$ Mpc.

The mapping between Eulerian real space $x(z)$ and redshift space $s(z)$ is given by: $s(z) = r(z) + v_r(z)$, with $v_r \equiv (v \cdot \hat{r})\hat{r}/(Ha)$; where $\hat{r}$ is the unit sight line vector, H the Hubble constant, a the scale factor, and $v = v(x)$ the 3D velocity field interpolated at the position of each halo in Eulerian-space r using the displacement field $\psi_{ALPT}(q)$. We split the peculiar velocity field into a coherent v^{coh} and a (quasi) virialized component v_σ: $v = v^{coh} + v^\sigma$. The coherent peculiar velocity field is computed in Lagrangian-space from the linear Gaussian field $\delta^{(1)}(q)$ using the ALPT formulation consistently with the displacement field (see equation 3):

$$v_{ALPT}^{coh}(q) = \mathcal{K}(q, r_S) * v_{2LPT}(q) + (1 - \mathcal{K}(q, r_S)) * v_{SC}(q), \tag{4}$$

with $v_{2LPT}(q)$ being the second order and $v_{SC}(q)$ being the spherical collapse component (for details see Kitaura, Yepes & Prada 2014). We use the high correlation between the local density field and the velocity dispersion to model the displacement due to (quasi) virialized motions. Effectively, we sample a Gaussian distribution function ($\mathcal{G}$) with a dispersion (see also Ata et al. 2017) given by $\sigma_v \propto (1 + \delta(r))^\gamma$. Consequently we assume,

$$v_r^\sigma \equiv (v^\sigma \cdot \hat{r})\hat{r}/(Ha) \frown \mathcal{G}(g \times (1 + \delta(r))^\gamma)\hat{r}. \tag{5}$$

For the Gaussian streaming model see Reid & White (2011); for non-Gaussian models see e.g. Tinker (2007). In closely virialized systems, the kinetic energy approximately equals the gravitational

MNRAS **500**, 3194–3212 (2021)

Downloaded from https://academic.oup.com/mnras/article/500/3/3194/5941513 by guest on 17 January 2025

potential and a Keplerian law predicts γ close to 0.5, leaving only the proportionality constant g as a free parameter in the model. We leave a detailed investigation of the impact of redshift space distortions for future work.

2.2 Probabilistic model

We use a Bayesian framework to draw samples of the density field $\delta(\boldsymbol{q})$ from a posterior probability density $\mathcal{P}(\delta(\boldsymbol{q}) \mid N_{\mathrm{G}}(\boldsymbol{q}), \boldsymbol{B}(\delta(\boldsymbol{q})), \boldsymbol{R}(\boldsymbol{q}))$. The posterior itself is a product of a prior and a likelihood function that we will describe in more detail in the following. We express the expectation value of galaxies per voxel $\lambda_i = \langle N_{\mathrm{G}}(\boldsymbol{q}) \rangle_i$ [2] for all voxels $i \in [1 \dots N_{\mathrm{C}}]$ as

$$\lambda(\boldsymbol{q})_i = \langle N_{\mathrm{G}}(\boldsymbol{q}) \rangle_i = f_{\bar{N}i}(\boldsymbol{q}) \sum_k R_{ik}(\boldsymbol{q}) B_k(\delta(\boldsymbol{q})), \qquad (6)$$

with the normalization factor $\boldsymbol{f}_{\bar{N}}(\boldsymbol{q}) = \bar{N}/\langle \boldsymbol{B}(\delta(\boldsymbol{q})) \rangle$ ensuring the right galaxy number density $\bar{N}$, as given by each survey. In our case, the response function $\boldsymbol{R}$ is limited to the completeness w_i in each cell i as described in Section 3.2: $R_{ij}(\boldsymbol{q}) = R_{ii}\delta_{ij}^{\mathrm{K}}$. We then relate the expected number counts of galaxies per voxel to the actually observed number counts N_{G} with a Poisson likelihood (Kitaura & Enßlin 2008; Kitaura et al. 2010) model:

$$\mathcal{L}(N_{\mathrm{G}} \mid \lambda(\boldsymbol{q})) = \prod_i \frac{\lambda_i(\boldsymbol{q})^{N_{\mathrm{G}i}} \exp(-\lambda_i(\boldsymbol{q}))}{N_{\mathrm{G}i}!}, \qquad (7)$$

where $\lambda(\boldsymbol{q}) = \lambda(f_{\bar{N}}(\boldsymbol{q}), \boldsymbol{B}(\delta(\boldsymbol{q})), \boldsymbol{R}(\boldsymbol{q}))$. This is an adequate assumption for tracers with a vanishing small-scale clustering, which otherwise become sources of super-Poissonity (e.g. Peebles 1980) that can be modelled with a negative-binomial likelihood (Kitaura et al. 2014; Neyrinck et al. 2014; Ata, Kitaura & Müller 2015). However, since the tracers are mapped to Lagrangian space at very high redshifts (e.g. $z = 100$), a deviation from Poissonity becomes insignificant except for the most massive galaxies (Modi, Castorina & Seljak 2017; Abidi & Baldauf 2018; Schmittfull et al. 2019), which is not the case for our galaxy samples (Laigle et al. 2016; see Section 3).

The above-mentioned conditions are ideal to describe the matter distribution with a lognormal prior towards high redshifts (Coles & Jones 1991). Its derivation is precisely based on a comoving framework at initial cosmic times, before shell crossing occurs. We apply a logarithmic transformation, which further linearizes the density field (Neyrinck, Szapudi & Szalay 2009) as

$$\delta_{\mathrm{L}}(\boldsymbol{q}) = \log(1 + \delta(\boldsymbol{q})) - \mu, \qquad (8)$$

with $\mu = \langle \log(1 + \delta) \rangle = -\log(\langle e^{\delta_{\mathrm{L}}} \rangle)$ (Kitaura et al. 2012a). Since we consider early cosmic times, the overdensity field has little power $|\delta| \ll 1$. Hence, the logarithmic transformation ensures positive densities ρ (with $\delta = \rho/\bar{\rho} - 1 = \exp(\delta_{\mathrm{L}} + \mu) - 1$), while we can model the prior $\pi(\delta(\boldsymbol{q}))$ [3] for the linear density field δ_{L} by a Gaussian distribution with zero mean:

$$\pi(\delta(\boldsymbol{q}) \mid \boldsymbol{C}_{\mathrm{L}}(\boldsymbol{q})) =$$
$$\frac{1}{\sqrt{(2\pi)^{N_{\mathrm{C}}} \det(\boldsymbol{C}_{\mathrm{L}}(\boldsymbol{q}))}} \exp\left(-\frac{1}{2}\delta_{\mathrm{L}}^{\dagger}(\boldsymbol{q}) \boldsymbol{C}_{\mathrm{L}}^{-1}(\boldsymbol{q}) \delta_{\mathrm{L}}(\boldsymbol{q})\right), \qquad (9)$$

where $\boldsymbol{C}_{\mathrm{L}}(\boldsymbol{q}) = \langle \delta_{\mathrm{L}}^{\dagger}(\boldsymbol{q}) \delta_{\mathrm{L}}(\boldsymbol{q}) \rangle$ is the covariance matrix of the linearized density fields, which depends on the cosmological parameters

---

[2]The expectation value is given by the ensemble average: $\langle X \rangle$.
[3]Note that the prior is actually a function of the linearized density field $\delta_{\mathrm{L}}(\boldsymbol{q})$, which is in turn a function of the original density field $\delta(\boldsymbol{q})$.

$\{p_{\mathrm{c}}\}$. Finally, we can express the posterior of $\delta(\boldsymbol{q})$ through Bayes theorem as

$$\mathcal{P}(\delta(\boldsymbol{q}) \mid N_{\mathrm{G}}(\boldsymbol{q}), \boldsymbol{B}(\delta(\boldsymbol{q})), \boldsymbol{R}(\boldsymbol{q}), \{p_{\mathrm{c}}\})$$
$$\propto \pi(\delta(\boldsymbol{q}) \mid \boldsymbol{C}_{\mathrm{L}}(\{p_{\mathrm{c}}\})) \times \mathcal{L}(N_{\mathrm{G}}((\boldsymbol{q})) \mid \lambda(\boldsymbol{q}), \boldsymbol{R}(\boldsymbol{q})), \qquad (10)$$

where the normalization given by the evidence is not necessary within HMC. Table 1 summarizes the main quantities that are sampled and how they are connected to each other.

2.2.1 Galaxy bias description

Finally, we need to specify the connection between the likelihood and the prior through the bias relation $\boldsymbol{B}(\delta)$. In the COSMIC BIRTH framework, non-local bias is described through the displacement. The split-background bias (Kaiser 1984; Bardeen et al. 1986), which is necessary in Eulerian space particularly for massive galaxies (see e.g. Kitaura et al. 2015), however, becomes negligible when homogenizing the galaxy distribution mapping it to Lagrangian space. Hence, we can assume a power-law Lagrangian bias as discussed in Kitaura et al. (2019):

$$\boldsymbol{B}(\delta(\boldsymbol{q})) = (1 + \delta(\boldsymbol{q}))^{b(z_q) f_b(z_q)}, \qquad (11)$$

where z_q is the redshift at which the Lagrangian coordinates are evaluated (for this study $z_q = 100$), b the linear large-scale bias, and f_{b} the non-linear correction factor of our bias description (Ata et al. 2017). The correction factor $f_b(z_q)$ can be determined iteratively as presented in Kitaura et al. (2019), ensuring that b exactly corresponds the large-scale linear bias. By using a power-law bias, we ensure that the density field is positive, since otherwise any bias less than one can potentially cause negative densities at voxels with δ close to -1.

The advantage of this bias description is that the only free parameter of our method is reduced to the large-scale bias at Eulerian space, which can be connected to Lagrangian space through passive evolution (Nusser & Davis 1994; Fry 1996):

$$b(z_q) = (b(z) - 1)\frac{D(z)}{D(z_q)} + 1, \qquad (12)$$

including the linear growth factor $D(z)$. We will show in Section 4.2 how we describe the large-scale bias evolution for the employed galaxy catalogues.

2.3 Summary of the algorithm

In summary, the joint probability distributions of all the above-mentioned variables can be expressed within a Gibbs-sampling scheme based on the corresponding conditional probabilities:

$$\delta(\boldsymbol{q}) \curvearrowleft \mathcal{P}_{\delta}(\delta(\boldsymbol{q}) \mid N_{\mathrm{G}}(\boldsymbol{q}), \boldsymbol{B}(\delta(\boldsymbol{q})), \boldsymbol{R}(\boldsymbol{q}), \boldsymbol{C}_{\mathrm{L}}(\boldsymbol{q}))$$
$$\{\boldsymbol{r}\} \curvearrowleft \mathcal{P}_r(\{\boldsymbol{r}\} \mid \{\boldsymbol{s}^{\mathrm{obs}}\}, \boldsymbol{v}_{\mathrm{r}}(\boldsymbol{q}), \mathcal{M}_v),$$
$$\{\boldsymbol{q}\} \curvearrowleft \mathcal{P}_q(\{\boldsymbol{q}\} \mid \{\boldsymbol{r}\}, \boldsymbol{\psi}(\boldsymbol{q}), \mathcal{M}_{\psi}),$$
$$\boldsymbol{R}(\boldsymbol{q}) \curvearrowleft \mathcal{P}_R(\boldsymbol{R}(\boldsymbol{q}) \mid \boldsymbol{R}(\boldsymbol{s}), \boldsymbol{\psi}(\boldsymbol{q}), \mathcal{M}_{\psi}),$$
$$\boldsymbol{B}(\delta(\boldsymbol{q})) \curvearrowleft \mathcal{P}_B(\boldsymbol{B}(\delta(\boldsymbol{q})) \mid \boldsymbol{B}(\delta(\boldsymbol{s})), \boldsymbol{\psi}(\boldsymbol{q}), \mathcal{M}_{\psi}), \qquad (13)$$

where the functional dependence of $\boldsymbol{q}$ and $\boldsymbol{s}$ stands for Lagrangian real-space and Eulerian redshift-space coordinates, respectively. The curved left arrows stand for the sampling process. $\mathcal{M}_v$ and $\mathcal{M}_{\psi}$ represent the models describing peculiar motions and displacement fields.

Table 1. Inferred quantities of COSMIC BIRTH: The Eulerian real-space positions $\{r\}$ are inferred from the observed redshift-space positions $\{s^{\rm obs}\}$ via the peculiar velocity $v_{\rm r}$. The Lagrangian positions $\{q\}$ are calculated from the Eulerian ones by applying the displacement field $\psi(q)$. The same mapping is used to calculate the response operator $R(q)$ in Lagrangian space. The expectation value of galaxy number counts λ is estimated from the galaxy counts $N_{\rm G}(q)$, connected by the normalization factor $f_{\bar{N}}(q)$, the bias model $B(\delta(q))$ and the response function $R(q)$). We sample the initial density density field $\delta(q)$ at Lagrangian coordinates q from the posterior probability function $\mathcal{P}(\delta(q) \mid N_{\rm G}(q), B(\delta(q)), R(q), C_{\rm L}(q))$. The connection quantities $v_{\rm r}, \psi(q), B(\delta(q)), R(q), C_{\rm L}(q)$ depend on a set of cosmological parameters $\{p_{\rm c}\}$. $f_{\rm b}$ is the bias correction term, derived from the large-scale linear bias b and a smoothing kernel $\mathcal{K}$ with radius $r_{\rm S}$. Note that, $\lambda(q), N_{\rm G}(q), B(\delta(q)), f_{\bar{N}}(q)$, and $\delta(q)$ are arrays of scalar quantities of $N_{\rm C}$ entries, while $v_{\rm r}(q), \psi(q)$ are arrays of 3D vector quantities of $N_{\rm C}$ entries. The quantities $R(q)$ and $C_{\rm L}(q)$ are matrix operators of $N_{\rm C} \times N_{\rm C}$ dimensionality.

Inferred quantity	Parent quantity	Connected via
$\{r\}$ Eulerian real-space	$\{s^{\rm obs}\}$ Eulerian redshift-space	$v_{\rm r}(q)$ Peculiar velocity
$\{q\}$ Lagrangian real-space	$\{r\}$ Eulerian real-space	$\psi(q)$ Displacement field
$R(q)$ Lagrangian response function	$R(s)$ Eulerian response function	$\psi(q), v_{\rm r}(q)$
$\lambda(q)$ Galaxy number expectation	$N_{\rm G}(q)$ Galaxy number counts	$f_{\bar{N}}(q), B(\delta(q)), R(q)$ (see caption)
$f_{\rm b}(z)$ Bias correction	$b(z)$ Linear bias	$B(\delta(q)), N_{\rm G}(q), \mathcal{K}(r_{\rm S})$ (see caption)
$\delta(q)$ Lagrangian density	$\{q\}$ Lagrangian real-space	$\mathcal{P}(\delta(q) \mid N_{\rm G}(q), B(\delta(q)), R(q), C_{\rm L}(q))$ (see caption)

2.4 Multitracer formalism in Lagrangian space

In this study, we aim at combining the data of five spectroscopic surveys in the COSMOS field that share spatially overlapping footprints and similar redshift distributions, however relying on different observing strategies (for more details see Section 3). Thus merging the surveys into a single catalogue by a pre-processing step, i.e. adding the different catalogues into one data set, represents a difficult task, which in general cannot be accomplished without making a series of simplifying assumptions, e.g. neglecting the different selection criteria.

Some of the previous pioneering Bayesian inference studies have already applied multitracer treatment, however, all performed in Eulerian space, and without taking into account separate survey footprints within the Bayesian scheme (Granett et al. 2015; Jasche, Leclercq & Wandelt 2015), or separate footprints from the same (super) set of catalogues (Jasche & Lavaux 2017; Jasche & Lavaux 2019).

In this work, we aim at performing joint analysis of entirely different surveys within the Bayesian framework. According to Section 2.2, we perform the reconstructions in Lagrangian space (in our case corresponding to a redshift of $z = 100$), where gravitational interactions have not yet introduced mode couplings of density perturbations. Therefore, mapping the galaxies to Lagrangian space before the density sampling step (see Table 1) is gradually reducing the covariance of the different surveys and homogenizes the galaxy fields. This enables us to treat each survey as a distinct component, avoiding mixed terms in the likelihood expression. Therefore, spatial overlap in Lagrangian space does not represent a problem, as long as we make sure that no galaxy is multiply counted among the different surveys.

In this way, we are able to combine different catalogues (indexed with superscript k) taking into account their distinct survey selection functions $R^k(q)$, number densities $N_{\rm G}^k(q)$, and galaxy bias functions $B^k(\delta(q))$.

The Eulerian to Lagrangian mapping of COSMIC BIRTH, shown in equation (2), can lead to a change of the redshift bin of a tracer. Consequently, the bias of this tracer will not coincide with the bias of the redshift bin at its new location. This can be taken into account by keeping track of galaxies staying at a redshift bin, or jumping from one redshift bin to another, which causes a 'bias mixing' implemented in the COSMIC BIRTH code (see section 3 in Kitaura et al. 2019). While this effect is negligible when interpolating the bias within the redshift bins, it has the advantage that a multitracer

treatment is already implemented in this framework. We can then extend COSMIC BIRTH to perform a full Bayesian multitracer & multisurvey analysis to address the challenges of this work following the calculations presented in, e.g. Ata et al. (2015, Appendix A) to express the corresponding posterior $\mathcal{P}^{\rm multi}$ in Lagrangian coordinates and combine the surveys with their specific likelihoods to construct the combined likelihood $\mathcal{L}^{\rm multi}$ by

$$\mathcal{L}^{\rm multi}\left(N_{\rm G}(q)|\lambda\left(f_{\bar{N}}(q), B(\delta(q)), R(q)\right)\right) \propto$$
$$\prod_k \mathcal{L}^{(k)}\left(N_{\rm G}^{(k)}(q)|\lambda^{(k)}\left(f_{\bar{N}}^{(k)}(q), B^{(k)}(\delta(q)), R^{(k)}(q)\right)\right), \quad (14)$$

where index $k \in [1 \dots N_{\rm S}]$ denotes the different surveys. Accordingly, we need to generalize equation (6) and distinguish between the surveys in the reconstructed volume:

$$\lambda^{(k)} = \left\langle N_{\rm G}^{(k)}(q)\right\rangle = f_{\bar{N}}^{(k)}(q)R^{(k)}(q)B^{(k)}(\delta(q)). \quad (15)$$

In Hamiltonian sampling, one seeks to draw samples of the potential energy term $\mathcal{U}$ of the Hamiltonian, that is linked to the posterior in equation (10) via

$$\mathcal{U} = -\ln \mathcal{P}. \quad (16)$$

Thus, the multitracer & multisurvey posterior $\mathcal{P}^{\rm multi}$ writes for a number of $N_{\rm S}$ surveys as

$$-\ln \mathcal{P}^{\rm multi}\left(\delta(q)|\lambda^{(1)}(q), \lambda^{(2)}(q), \dots, \lambda^{(N_s)}(q)\right) =$$
$$c - \ln \pi(\delta(q) \mid C_{\rm L}(q))$$
$$-\ln \mathcal{L}^{(1)}\left(N_{\rm G}^{(1)}|\lambda^{(1)}\left(f_{\bar{N}}^{(1)}(q), B^{(1)}(\delta(q)), R^{(1)}(q)\right)\right)$$
$$-\ln \mathcal{L}^{(2)}\left(N_{\rm G}^{(2)}|\lambda^{(2)}\left(f_{\bar{N}}^{(2)}(q), B^{(2)}(\delta(q)), R^{(2)}(q)\right)\right)$$
$$\vdots$$
$$-\ln \mathcal{L}^{(N_{\rm S})}\left(N_{\rm G}^{(N_{\rm S})}|\lambda^{(N_{\rm S})}\left(f_{\bar{N}}^{(N_{\rm S})}(q), B^{(N_{\rm S})}(\delta(q)), R^{(N_{\rm S})}(q)\right)\right), \quad (17)$$

where the constant c does not depend on $\delta(q)$.

3 SURVEY DATA AND COMPLETENESS

Within the COSMOS field, several spectroscopic surveys have been undertaken (see e.g. Hasinger et al. 2018, for a summary), focusing mainly on star-forming galaxies at high redshifts. This work uses data from five different surveys, summarized in Table 2. We describe each survey in more detail below and afterwards explain our method to estimate the corresponding survey completenesses.

Table 2. Summary of five surveys used for this study. N^{Obj} resembles the number of galaxies that we use after applying spectroscopic quality criteria and removing duplicates. (∗) zCOSMOS-deep, VUDS, and MOSDEF exceed the redshift range of our reconstructions. MOSDEF observes the redshift range in intervals of $1.37 \leq z \leq 1.70$, $2.09 \leq z \leq 2.61$, and $2.95 \leq z \leq 3.80$. In the case of FMOS–COSMOS, about half of the galaxies in the central footprint are used in this work.

Survey	N^{Obj}	z range	Parent catalogue
zCOSMOS-deep	3544	$1.4 \leq z \leq 3.6$(∗)	COSMOS
VUDS	1822	$1.4 \leq z \leq 3.6$(∗)	COSMOS
MOSDEF	401	$1.4 \leq z \leq 3.6$(∗)	3D-*HST*
ZFIRE	149	$2.0 \leq z \leq 2.2$	ZFOURGE
FMOS–COSMOS	587	$1.4 \leq z \leq 1.7$	COSMOS

3.1 Surveys in the COSMOS field

Let us briefly recap the main characteristics of the different spectroscopic galaxy surveys considered in this work.

(i) zCOSMOS-deep

The zCOSMOS-deep survey is the high-redshift component of the zCOSMOS spectroscopic survey, which covers the central 1 deg^2 of the zCOSMOS footprint (Lilly et al. 2007; Lilly et al. in preparation) using the VIMOS spectrograph (LeFevre et al. 2003) on the VLT. The targets were chosen from the then-current version of the multicolour photometric COSMOS catalogue (Capak et al. 2007). To isolate galaxies at redshifts of $z > 1.5$, several selection criteria were applied. In particular, a selection in the $(U - B)/(V - R)$ colour–colour plane, called 'UBR' selection (Steidel et al. 2004) was combined with the 'BzK' selection (Daddi et al. 2004). For both the 'UGR' and 'BzK' selections, an additional selection of $22.5 < B_{\text{AB}} < 25.0$ and a deep K-band imaging reaching down to $K_{\text{AB}} \sim 23.5$ were applied. Our analysis is done on a tentative version of the zCOSMOS-deep catalogue that has been used for the bulk of previous works that employ zCOSMOS-deep data. A refined version of this catalogue will be available in the future (Lilly et al., in preparation).

(ii) VUDS

The VIMOS Ultra Deep Survey (Le Fèvre et al. 2015), hereafter VUDS, is another spectroscopic survey carried out on the VIMOS spectrograph that was partly operated in the COSMOS field, but considerably deeper than zCOSMOS-deep with up to $\sim 3 \times$ larger integration times. VUDS peaks in number density at $z \sim 3$ and was designed to explore multiple questions, including those related to the formation rates of stars and merging of galaxies during the period of time when galaxies were most active. Another major success was to identify and characterize galaxy protoclusters at $z \sim 2$–4 (e.g. Cucciati et al. 2014, 2018; Lemaux et al. 2018). The selection is based on photometric redshift selections (Ilbert et al. 2013), with a small fraction of galaxies selected using the Lyman-break technique (Steidel et al. 1996). VUDS and zCOSMOS are based on different versions of COSMOS parent catalogue, so the astrometry was not identical for all sources between the two surveys. Because of the different catalogues used, we employed a matching radius of 0.1–0.2 arcsec to identify true duplicates.

(iii) MOSDEF

The MOSFIRE Deep Evolution Field (MOSDEF) Survey (Kriek et al. 2015) is partly taken in the COSMOS field, separated into three redshift intervals at $1.37 \leq z \leq 1.70$, $2.09 \leq z \leq 2.61$, and $2.95 \leq z \leq 3.80$, down to fixed H_{AB} magnitudes of 24.0, 24.5, and 25.0 for each interval. The MOSFIRE spectrograph (McLean et al. 2012) on the Keck-I telescope was used to obtain near infrared emission
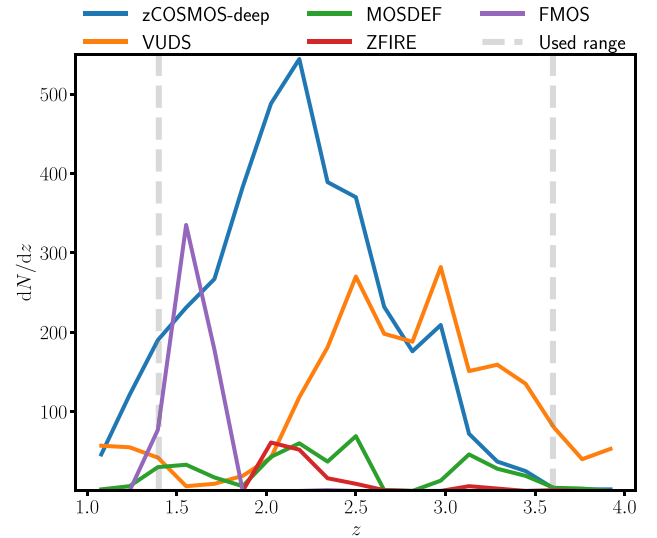

Figure 1. Radial distribution functions for the five considered surveys shown as function of redshift with a bin width of $\Delta z = 0.16$. The two grey-dashed vertical lines signpost the considered redshift range $1.4 \leq z \leq 3.6$ in this work.

line redshifts of the targeted galaxies. The targets were selected from the photometric and grism 3D-*HST* (Brammer et al. 2012) data, applying photometric redshift and magnitude requirements to the parent catalogue and show a spectroscopic success rate of around ~ 80 per cent.

(iv) ZFIRE

The KECK/MOSFIRE Spectroscopic Survey of Galaxies in Rich Environments at $z \sim 2$ (ZFIRE; Nanayakkara et al. 2016) using the MOSFIRE spectrograph was partly taken in the COSMOS and Hubble Ultra Deep Survey field (Beckwith et al. 2006). For the COSMOS field, the targets were K band selected from the photometric parent FourStar Galaxy Evolution Survey (ZFOURGE) catalogue (Straatman et al. 2016), requiring a photometric redshift $2.0 \lesssim z_{\text{phot}} \lesssim 2.2$. ZFIRE was designed to observe primarily the $z = 2.095$ galaxy cluster (Yuan et al. 2014).

(v) FMOS–COSMOS

The FMOS–COSMOS survey (e.g. Silverman et al. 2015; Kashino et al. 2019; hereafter FMOS) used the Fibre MultiObject Spectrograph (Kimura et al. 2010) at the Subaru Telescope, observing star-forming galaxies at redshifts $z \sim 1.6$ in the near infrared. The targets were selected from the COSMOS photometric catalogue within a redshift range of $1.4 \lesssim z_{\text{phot}} \lesssim 1.7$, additionally applying limits on the UltraVISTA K-band magnitude limit and the Hα flux predicted from SED fitting. FMOS objects are pre-selected with secure photometric redshift using the filters available in the COSMOS photometric survey. In this analysis, we utilize FMOS observations in the range of $149.8 \leq \text{RA} \leq 150.4$ and $1.8 \leq \text{Dec.} \leq 2.5$.

In summary, the surveys apply different colour, magnitude, and photometric redshift pre-selection cuts to the parent photometric catalogues to efficiently select high redshift targets for spectroscopy. In the case of galaxies that were spectroscopically observed in more than one of these surveys, we kept the one with better redshift quality, based on the signal-to-noise ratio of the observed spectra (see Section 3.2.1). From the redshift distributions of the surveys shown in Fig. 1, we can see that the number density of galaxies
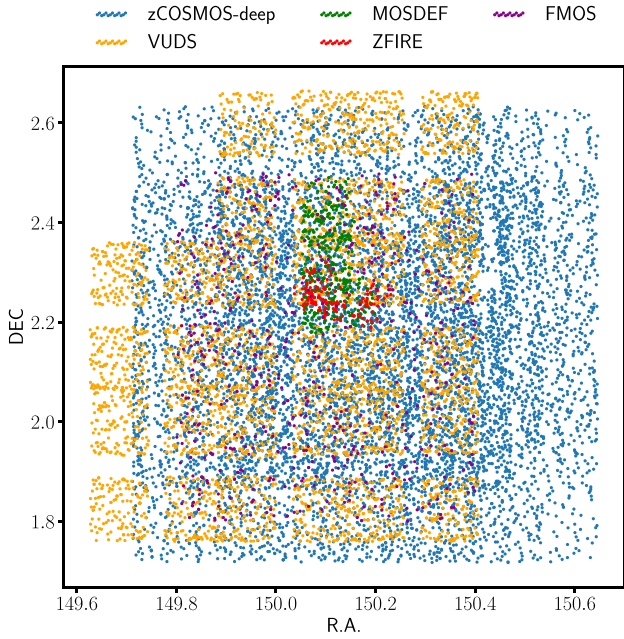
Figure 2. Survey footprints: Spatial distribution of the observed galaxies in the right ascension (RA)–declination (Dec.) plane. We follow the same colour coding as in Fig. 1. At a redshift of $z = 2.5$, the maximum separation $\Delta\mathrm{Dec.} = 0.91°$ corresponds to a comoving transverse distance of $d_{\Delta\mathrm{DEC}} = 64.8\,h^{-1}$ Mpc, while for $\Delta\mathrm{RA} = 1.02°$ the maximum separation is $d_{\Delta\mathrm{RA}} = 70.2\,h^{-1}$ Mpc assuming the cosmological parameters in Section 1.

peaks at $2 \leq z \leq 2.5$, mainly contributed by the zCOSMOS-deep survey.

The footprints (galaxy distribution in angular coordinates) of each survey are shown in Fig. 2. From this figure, we can recognize that the edges of zCOSMOS-deep, represented in blue, have been observed with only a single pointing, thus having a lower number density, while in the central area the pointings are overlapping, allowing for a denser targeting. Regarding the VUDS survey, represented in orange, Fig. 2 shows the gaps in between the quadrants of an individual VIMOS pointing, whereas within one quadrant the target sampling is very homogeneous. Focusing on the MOSDEF and ZFIRE surveys depicted in green and red, respectively, one can see that they cover smaller areas in the central part of the COSMOS field. The FMOS survey, represented in magenta, shows a sparser targeting compared to the rest of the surveys considered in this study. While the three surveys MOSDEF, ZFIRE, and FMOS lie withing the footprints of zCOSMOS-deep and VUDS, the latter ones only partly overlap. VUDS shares 86 per cent angular coverage with zCOSMOS-deep, which corresponds to 69 per cent VUDS coverage of the zCOSMOS-deep footprint.

3.2 Survey completeness estimation

The response function $\boldsymbol{R}$ (see Table 1, Section 2) represents the efficiency how each voxel i has been observed as compared to the number of possible targets (see Fig. 3 for details of the selection stages). In a Bayesian analysis, the likelihood function should therefore account for the uncertainty of the galaxy number counts $N_\mathrm{G}(\boldsymbol{q})$ as a function of the completeness, shown in equation (6). The response function is calculated from an angular and a radial component, $\boldsymbol{R}_\alpha$ and $\boldsymbol{R}_r$, respectively, as they can be independently
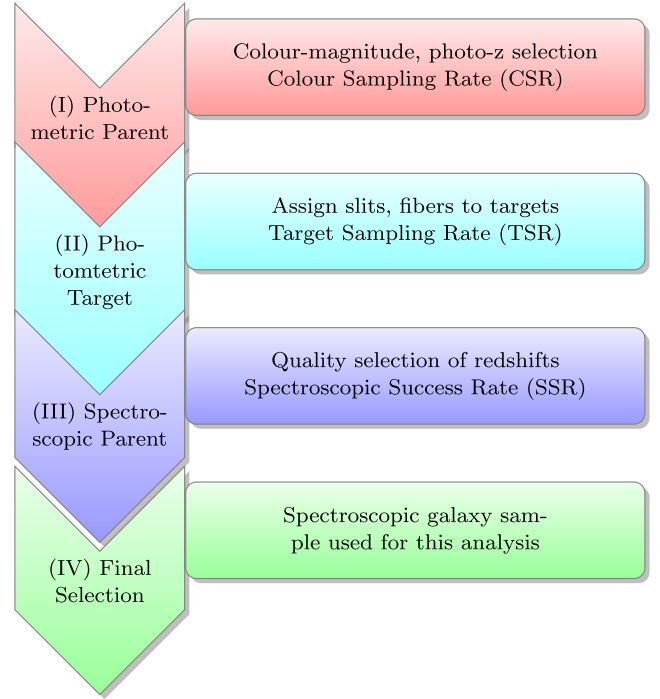

Figure 3. Flowchart describing the process to obtain the final spectroscopic catalogue, starting from the parent photometric survey. The arrow symbol on the left represents the name of the survey. The descriptive text on the right denotes the selection/operation that is undertaken upon the survey, leading to the next stage. We use an adapted terminology of the VVDS/VIMOS collaboration: Stage (I) to (II) is called colour sampling rate (CSR), (II) to (III) target sampling rate (TSR), and finally (III) to (IV) spectroscopic success rate (SSR).

calculated:

$$\boldsymbol{R} = \boldsymbol{R}_\alpha \cdot \boldsymbol{R}_r . \tag{18}$$

Both components are calculated on the reconstructed mesh grid and then multiplied for each voxel. It is practical to consider them separately, since the angular part is not subject to redshift space distortions, contrary to the radial part. Following Kitaura et al. (2019), we can compute the angular response operator in Eulerian space $\boldsymbol{R}_\alpha(\boldsymbol{s}) = \boldsymbol{R}_\alpha(\boldsymbol{r})$ once, and in each Gibbs-sampling iteration map it to Lagrangian space $\boldsymbol{R}_\alpha(\boldsymbol{q})$ through the displacement field (see Table 1). The radial response function can be trivially computed from the distribution of large-scale tracers in Lagrangian real space coordinates $\{\boldsymbol{q}\}$ in each Gibbs-sampling iteration, and multiplied according to equation (18). Let us thus focus in detail on the computation of the angular completeness.

3.2.1 Angular completeness

In the following, we adopt the terminology of the VIMOS related surveys (e.g. VVDS, VIPERS; see e.g. Ilbert et al. 2005; Zucca et al. 2009; Guzzo et al. 2014; Granett et al. 2015; Scodeggio et al. 2018) to estimate the spatial completeness in the reconstructed volume.

To compute the angular completeness, we need to distinguish between four selection stages, starting from the photometric parent catalogue and yielding the final spectroscopic survey, shown in Fig. 3. In the first stage, observers apply photometric selection criteria on top of the *parent photometric catalogue* (I) to select galaxies with certain properties (e.g. star forming, redshift range, etc). The resulting

catalogue consists of *photometric targets* (II), leading to a second stage, as shown in Fig. 3. The transition from the photometric parent (I) to the photometric target catalogue (II) is called *colour sampling rate* (CSR) and accounts for the colour–colour, colour-magnitude and photometric redshift selections. A fraction of the photometric targets (II) is chosen for spectroscopy, which we call *Spectroscopic Parent* (III). The ratio of (II) and (III) is called *Target Sampling Rate* (TSR). Many factors have an influence on the final selection of spectroscopic galaxies such as the spatial arrangement of the slits/fibres, the conditions during the observations etc. This means, that not all the spectra taken for individual galaxies can be translated into a reliable redshifts. We call the ratio of the spectroscopic parents (III) to the final selection (IV) the *spectroscopic success rate* (SSR).

The selection estimation approach consists on reproducing the targeting strategy of each survey to precisely estimate the ratio from the parent photometric catalogue to the final spectroscopic galaxies' selection. This has been neglected in previous studies based on the COSMOS field (see Section 5 and references therein).

Let us describe the different selection steps in more detail:

(i) **Colour sampling rate**
One can make the robust assumption that the photometric pre-selections applied on the parent photometric catalogue (I) are constant over the footprint of each survey. In such a case, the CSR will only result in an overall normalization factor of the number density of galaxies (see Pezzotta et al. 2017), but not influence the angular dependent clustering. Therefore, we can safely absorb this factor into the radial selection factor.

(ii) **Target sampling rate**
First, we reproduce the photometric pre-selection criteria. Then we build the ratio of the photometric targets N_{II} with the number of the spectroscopic parents N_{III}, regardless of the quality of the spectra, $TSR = N_{III}/N_{II}$. We construct a mesh grid 80×80 cells over the RA–Dec. plane (Fig. 2), resulting in a resolution of about 0.9 arcmin per cell. We note that the reconstructions are computed with a comoving resolution of $d_R = 2\,h^{-1}$ Mpc (see Section 5). This corresponds to a angular aperture of $\vartheta = 1.5$ arcmin at a redshift of $z = 3.6$ and thus coarser than the angular completeness resolution. We set the value of the selection function in between the VUDS quadrants and outside the borders of the surveys to zero.

(iii) **Spectroscopic success rate**
Finally, we select a subset N_{IV} of galaxies from the spectroscopic parent sample, that have high redshift accuracies, as described follows. For the zCOSMOS-deep and VUDS surveys we demand redshift flags of ≥ 2. For the MOSDEF survey we apply quality flags of $= 3$, which corresponds to a redshift confidence of > 95 per cent and a minimum signal-to-noise ratio of $2 \leq S/N \leq 3$. For the ZFIRE survey, we apply quality flag of 2, which corresponds to a $S/N \geq 5$ and $|z_{spec} - z_{phot}| \leq 0.2$. For the FMOS survey we demand $3 \leq S/N \leq 5$, which translates into a redshift quality flag of ≥ 2. The SSR is then calculated as $SSR = N_{IV}/N_{III}$.

The resulting angular selection masks

$$w_\alpha^k = TSR^k \times SSR^k, \qquad (19)$$

in the RA–Dec. plane are shown in Fig. 4 for each survey k. In the case of the zCOSMOS-deep survey the higher targeting rate in the center of the survey footprint, caused by several overlapping pointings with the VIMOS spectrograph, can be appreciated. For the VUDS survey on the contrary, each pointing is unique, and hence, the areas do not overlap. This results in thin stripes between the quadrants of the pointings in the footprint, showing the inter-CCD gaps of ~ 2 arcmin in the VIMOS focal plane.
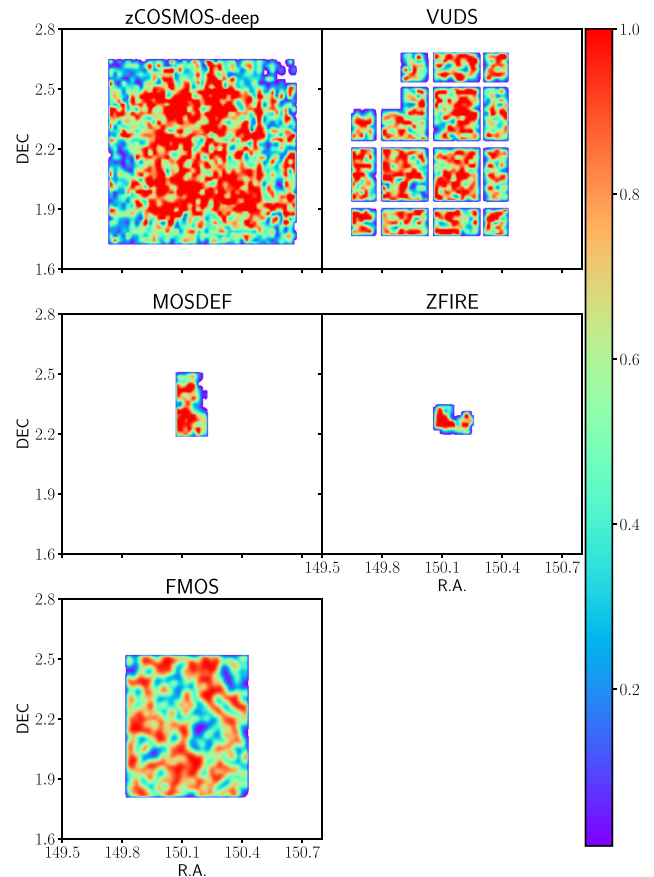

Figure 4. Final angular completeness mask w_α^k for all five surveys computed from TSR and SSR in the RA–Dec. plane normalized to unity.

Once we have computed the angular mask, we need to project it into three dimensions, as required by the COSMIC BIRTH code.

In particular, we project each RA–Dec. bin value of w_α^k into our cubical mesh grid on which we perform the reconstructions. As the resolution of w_α^k is higher compared to the reconstruction mesh grid, we average over each sight at a voxel i and thus obtain the angular completeness R_{ii}^k for all k surveys and cell i. Fig. 5 exemplary shows the projection of the VUDS angular selection mask into the reconstructed volume. The data preparation and details of the coordinate system are explained in Section 5.1. The top panel of Fig. 5 shows the stripes and the corresponding gaps in between the pointings of the survey in the Eulerian frame $R_\alpha(s)$, in accordance with Figs 2 and 4. The VUDS inter-CCD gaps translate into empty regions of $2.3\,h^{-1}$ Mpc at redshifts of $z = 3.6$. The translation of the angular response operator to Lagrangian space $R_\alpha(q)$ through the action of gravity causes a deformation in the survey window, which is represented in the lower panel of Fig. 5. This means, that unobserved angular regions in Eulerian space, might have been effectively partially observed in Lagrangian space. In turn, the deformation of the survey window can also cause that regions are effectively unobserved in Lagrangian space, that were observed in Eulerian space (see Fig. 5, e.g. around redshift $z \sim 2$). As expected, we can see that the deformation is stronger towards lower redshifts, where the growth of structures is more evolved.
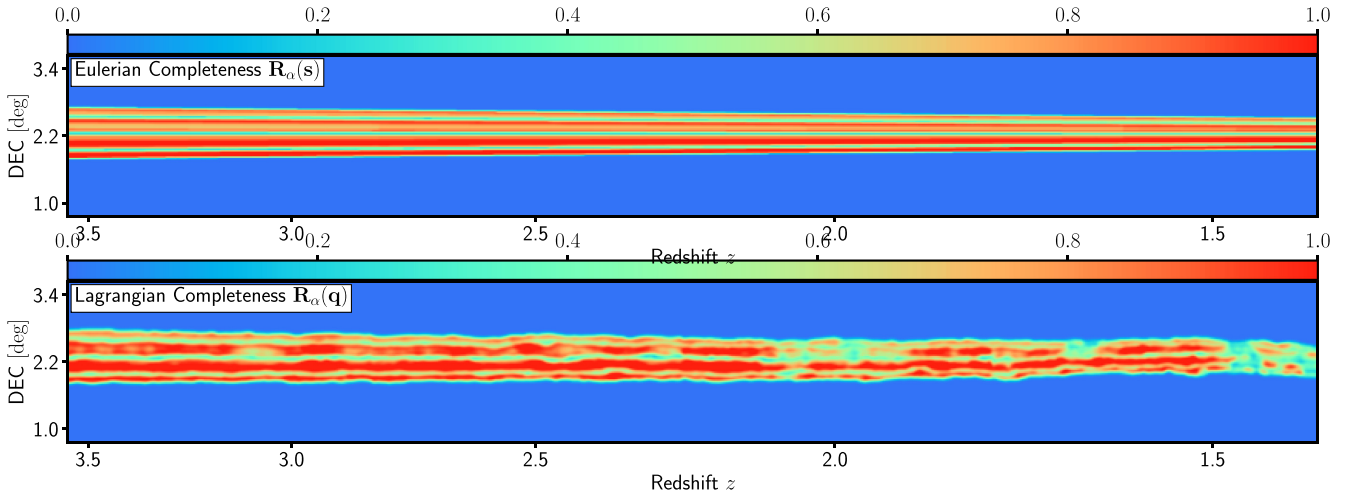
Figure 5. Single slice plot of the projected window function $\mathbf{R}_\alpha$ for the VUDS survey into the reconstructed volume, normalized to unity, showing 0 for non-observed regions and 1 for maximum completeness. *The top panel* represents the window calculated directly from the observations in Eulerian space s, showing the effects of the gaps in between the VUDS quadrants (compare Fig. 4), emphasizing the importance of an accurate selection function handling for this work. *The bottom panel* presents the same slice in Lagrangian space q. The whole declination angle on the Y-axis corresponds to $200\,h^{-1}$ Mpc in comoving distance.

4 LARGE-SCALE GALAXY BIAS

The COSMIC BIRTH algorithm accounts for stochastic and non-linear Lagrangian bias, while non-local Eulerian bias is modelled through the displacement field connecting Eulerian to Lagrangian space (see Section 2 and for further details in Kitaura et al. 2019). The only free parameter is the large-scale Eulerian bias (Kaiser 1984), which needs to be determined from observations or simulations. We rely on detailed bias studies of highly star-forming galaxies due to the nature of the surveys considered in this study (see Section 4.1).

4.1 Bias studies in simulations and observations

The various galaxy properties are in general correlated with their clustering behaviour, and hence, are indicators of how they trace the underlying dark matter density field. More massive and luminous galaxies, such as luminous red galaxies (LRGs), for instance (see e.g. Alam et al. 2017), show a strong clustering, tracing mainly the peaks of the density field (Kitaura et al. 2015). These galaxies are passively evolving, showing low stellar formation activity and old stellar populations.

Apart from LRGs, star-forming galaxies can be identified with photometric techniques (see e.g. Daddi et al. 2004) and emission line spectroscopy, frequently called emission line galaxies (ELGs) in literature. The galaxy bias of highly star-forming galaxies, such as [O II], [O III], H α-detected galaxies (see e.g. Delubac et al. 2017; Kaasinen et al. 2017), UV emitting Lyman-break galaxies (Kollmeier et al. 2003) and Ly α emitters (LAEs, see e.g. Kennicutt (1998)) have been extensively studied in the literature.

Unlike LRGs that trace only the densest peaks of the density field, ELGs can populate the density field at nearly the whole range of overdensities. Therefore, their bias is close to unity at low redshifts (Favole et al. 2016). However, in this work we consider higher redshifts, when the dark matter field was less evolved, displaying weaker density perturbations. Hence, these types of galaxies display an increasing bias towards high redshifts (Okada et al. 2016; Guo et al. 2019).

Studies based on numerical simulations have shown the bias as a function of redshift and star formation rate, comparing the galaxy and matter density fields (see table 1 in Chiang, Overzier & Gebhardt 2013). Within the VUDS survey a bias measurement for the redshift range of $2 \leq z \leq 5$ has been accomplished in Durkalec et al. (2015, 2018) using two-point clustering analysis. Similar analysis has been performed for the FMOS–COSMOS survey, studying the projected correlation function and from there estimating the bias at a median redshift of $\bar{z} = 1.58$ (Kashino et al. 2017).

4.2 Large-scale bias evolution

In the following, we explain our strategy to compute the large-scale bias, as required to perform the dark matter reconstructions throughout the redshift range of $1.4 \leq z \leq 3.6$.

Given that the galaxy populations from the five considered surveys are similar, we assume as a null hypothesis, that galaxies share the same bias at a given redshift, and that their bias passively evolves through equation (12; equivalent to a perfect correlation $r(z_2) = 1$ in equation 20). Then, we choose a narrow redshift range for which the deviation from passive evolution is expected to be negligible. In particular, we select the range $1.4 \lesssim z \lesssim 1.8$ embedded in a rectangular volume that extends from $2875\,h^{-1}$ Mpc $\leq d_L \leq 3387\,h^{-1}$ Mpc in line-of-sight distance. We perform a series of COSMIC BIRTH runs with the data set in this volume with a varying bias at $z = 1.5$ from 1 to 2.5 (the general set-up of the COSMIC BIRTH reconstructions is described in Section 5.1). From these runs, we find that the reconstructed primordial matter density shows unbiased power spectra w.r.t. the theoretical linear one for $b \approx 1.9 \pm 0.3$ at redshift $z = 1.5$ being conservative (and the rest of bias values given by passive evolution). This is in good agreement with the findings of Cochrane et al. (2017, see table 4, where they obtain $b = 1.78^{+0.06}_{-0.08}$ at $z = 1.48$), and also roughly compatible with the clustering analysis of the FMOS–COSMOS survey (Kashino et al. 2017, see table 3, finding $b = 2.44^{0.38}_{0.32}$ at median redshift $\bar{z} = 1.59$) within the estimated uncertainties.

The passive evolution model (Nusser & Davis 1994; Fry 1996), based on our low redshift bias measurement, evolves to higher red-
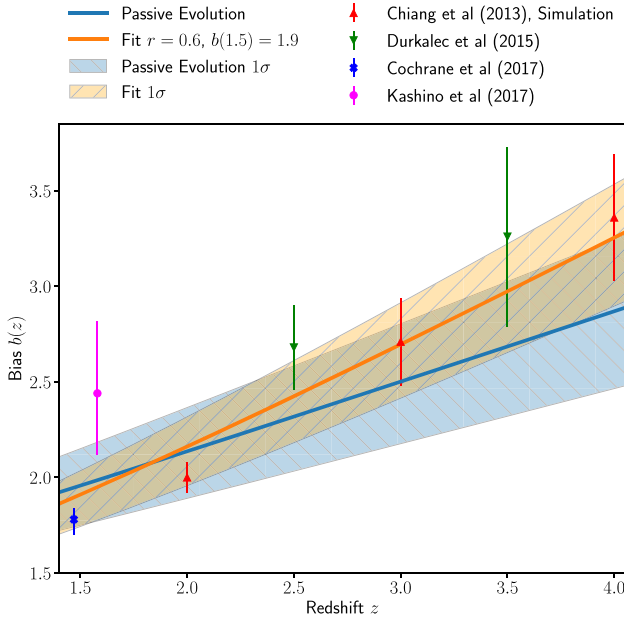
Figure 6. Large-scale galaxy bias $b(z)$ as a function of redshift z. The blue solid line and the corresponding blue-shaded error band show the best fit for the passive evolution model of equation (12) with $b(z = 1.5) = 1.9 \pm 0.33$. The solid orange line and the orange-shaded error band show the fitting result assuming equation (20) with the parameters $r(z = 1.5) = 0.6$ and $b(z = 1.5) = 1.9 \pm 0.33$. The circles denote previous bias measurements, whereas the red triangles show values derived from numerical simulations.

shifts according to the blue solid line including the large associated uncertainties represented by the shaded light blue area in Fig. 6. From this figure we can assume that there is a trend in the bias measurements towards higher bias values with increasing redshift than what is predicted by passive evolution, although within the large uncertainties the passive evolution model is still nearly compatible with the data points.

In such a case the redshift evolution of the galaxy populations can be modelled. This can be achieved by introducing a correlation coefficient r between the dark matter and the galaxy density field as a function of redshift (Tegmark & Peebles 1998). As the selection criteria of the surveys do not significantly change with observed distance, we expect a moderate enhancement of the galaxy bias towards higher redshifts beyond the one described by passive evolution. Thus, we write:

$$b(z_2) = \frac{\sqrt{\left(1 - \frac{D(z_2)}{D(z_1)}\right)^2 - 2r(z_1)\left(1 - \frac{D(z_2)}{D(z_1)}\right)b(z_1) + b^2(z_1)}}{\left(\frac{D(z_2)}{D(z_1)}\right)} \quad (20)$$

for two redshifts z_1 and z_2 with $z_1 < z_2$, a correlation coefficient $r(z_1)$, and the linear growth function $D(z)$ given by

$$D(z) = \frac{H(z)}{H_0} \int_z^\infty \frac{dz'}{H^3(z')} \Big/ \int_0^\infty \frac{dz'}{H^3(z')}, \quad (21)$$

normalized to unity at redshift zero $D(z = 0) = 1$. The evolution of the bias and the correlation coefficient are coupled as:

$$r(z_1) = \left(\left(1 - \left(\frac{D(z_2)}{D(z_1)}\right)\right) + r(z_2)\left(\frac{D(z_2)}{D(z_1)}\right)b(z_2)\right) \Big/ b(z_1). \quad (22)$$

A perfect correlation of $r(z_2) = 1$ at an earlier redshift z_2 will remain like that for all times, whereas $r(z_1)$ always tends towards 1, regardless of its initial value, according to equation (22). In the case of a perfect correlation $r(z_2) = 1$, equation (20) equals equation (12) (described in Section 2) and no change of the galaxy population is expected along redshift. However, $r(z_2) < 1$ implies a varying correlation coefficient, effectively describing a cosmic evolution of the galaxy distribution, which may be caused by galaxy formation or evolution (Tegmark & Peebles 1998).

On the other hand, a series of COSMIC BIRTH runs disfavour 2σ deviations from the upper bias limits quoted in the literature, as they lead to unreasonable biased dark matter reconstructions (see Appendix A). We note, that the selection function can lead to an excess of power in the two point statistics on large scales, and thereby higher bias values can be inferred (see e.g. Thomas, Abdalla & Lahav 2011). We therefore investigate the theoretical predictions for the bias evolution in simulations. As a result, we find that the galaxy samples considered in this work cover the stellar mass range of $M_* = 10^{9.5}\,M_\odot$ to $M_* = 10^{10.5}\,M_\odot$, peaking at $M_* \sim 10^{9.8}\,M_\odot$ (Lamaux et al., in preparation). According to this finding, we obtain the data points represented in red upwards pointing triangles in Fig. 6 (see table 1 in Chiang et al. 2013). We find that these simulation based data are in agreement with the observational measurements, however favour slightly lower bias values. In the spirit of being conservative, and avoid bias ranges which can be affected by selection effects, we include the simulation (Chiang et al. 2013) and observational (Durkalec et al. 2015, 2018; Cochrane et al. 2017; Kashino et al. 2017) data points in a least squares fit. The resulting bias evolution model is represented in solid red with the uncertainty given by the light red shaded area in Fig. 6.

According to this bias study, we have found some moderate evidence (given the hitherto large uncertainties due to the small volumes covered by high redshift galaxy surveys) for a cosmic evolution of the galaxy bias beyond passive evolution, hinting towards ongoing galaxy formation and merging processes at these redshifts. A coefficient of $r(z = 1.5) = 0.6$ corresponds to $r(z = 3.6) = 0.42$ (see equation 22). This apparently tiny variation, reduces the tension with observationally constrained biases at redshift $z > 1.5$. However, a proper verification of a deviation from passive evolution requires a deeper study, extending the runs we performed at low redshift to higher ones. The available data at this stage might not be sufficient to make stronger claims and we leave a more detailed investigation of the bias evolution to a forthcoming work.

The resulting large-scale bias calculations presented in this section can be fed into the COSMIC BIRTH code to produce unbiased dark matter reconstructions, as shown in Section 5.

5 COSMIC BIRTH APPLIED TO THE COSMOS FIELD

In this section we present the application of COSMIC BIRTH code (see Kitaura et al. 2019, and Section 2) to the spectroscopic surveys in the COSMOS field (see Section 3). Previous pioneering COSMOS density field estimates were based on photometric redshifts (Kovac et al. 2010) and used tessellations of the observed galaxy fields (Scoville et al. 2013; Darvish et al. 2015), but did not account for the selection function (Amara et al. 2012; Smolčić et al. 2017), or focused on individual high density peaks (e.g. Wang et al. 2016). Therefore, this work represents the first comprehensive study to address all the above mentioned issues, taking into account structure formation, selection functions, redshift-dependent bias descriptions and redshift-space distortions within a forward Bayesian analysis.

5.1 Set-up of the reconstructions

The `COSMIC BIRTH` code in its current version performs calculations on cubical regular meshes in comoving Cartesian coordinates and uses the corresponding galaxy positions in redshift-space on the light-cone with their corresponding bias obtained from galaxy catalogues as input source.

First, we assume in this study a ΛCDM model with cosmological parameters defined in equation (1). Then we select the input catalogues comprehending five different redshift surveys, as described in Section 3.1. The large-scale bias as a function of redshift is obtained from the data itself and further constrained according to some previous studies, as explained in Section 4.2. From this, the connection to Lagrangian bias including a non-linear and non-local treatment is internally computed, as explained in Section 2.2.1. With the given cosmology, we can translate the angular and redshift coordinates for each galaxy into comoving Cartesian coordinates (x, y, z). The corresponding geometry and angular completeness to each survey is computed, as explained in detail in Section 3.2.1. While the `COSMIC BIRTH` code does not assume the plane-parallel approximation at any step, this approximation is nearly fulfilled given the large distances to the galaxies of the considered redshift range and the narrow angular coverage (see Fig. 1). We take advantage of that for visualisation purposes, and choose a coordinate transformation so that the centre of the zCOSMOS-deep survey is aligned to Dec. = 0° and RA = 180°, which makes the y-axis approximately coincide with the declination angle (for angles close to zero, as in this case) and the x-axis with the redshift z (see e.g. Figs 5 and 11). The final results are presented in the original coordinate system. The reconstructions comprise the galaxies of the five mentioned surveys within a redshift range of $1.4 \leq z \leq 3.6$, which translates into a comoving distance of $d_{Box} = 1898 \, h^{-1} \, \text{Mpc}$ along the line of sight. For computational reasons, we have split the reconstruction into four cubical volumes making sure that a large enough volume (in terms of mode-coupling) is taken in each case of $512 \, h^{-1} \, \text{Mpc}$ side length (Sorce et al. 2016). This resulted in a mesh grid resolution resolution of $2 \, h^{-1} \, \text{Mpc}$ using meshes of 256^3 voxels. We placed the four volumes successively along the line of sight considering an overlapping region of $50 \, h^{-1} \, \text{Mpc}$ on each side. This is an adequate choice, acknowledging that the galaxy–galaxy correlation function drops steeply at scales larger than $20 \, h^{-1} \, \text{Mpc}$ (see e.g. Anderson et al. 2012). However, boundary effects (e.g. velocity correlations) at the transition region of the sub-volumes are the primary source of uncertainty in the reconstructions, which will be further analysed in forthcoming works. Additionally, we did not place galaxies in a buffering zone of $25 \, h^{-1} \, \text{Mpc}$ at the edges of each volume, which we took into account within the radial selection function accordingly. This means that the line-of-sight data region is 5–7 times larger than in transverse directions for each sub-volume (see Fig. 2). Thus, the full reconstructions extends from $2875 \, h^{-1} \, \text{Mpc}$ to $4773 \, h^{-1} \, \text{Mpc}$ in comoving line-of-sight distance. We note, that `COSMIC BIRTH` code takes light-cone evolution into account within each reconstructed volume. We choose eight redshift bins for each sub-volume (see fig. 4 in Kitaura et al. 2019).

5.2 Numerical assessment and convergence

Our numerical tests showed that the `COSMIC BIRTH` reconstructions are insensitive in terms of the power spectrum to deviations from up to 30 per cent from our best large-scale bias evolution estimate shown in Fig. 6. We note, that this already excludes the high end of
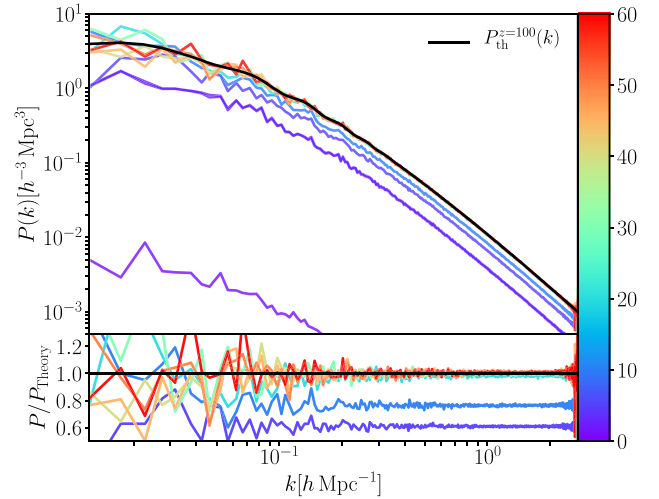
Figure 7. Power spectra of the inferred initial density field $\delta(q)$ at $z = 100$ shown colour coded for the first 60 HMC samples for the sub-volume ranging from $1.7 \leq z \leq 2.2$. In the bottom panel, the ratio of the power spectra and the theoretical prediction is shown. Convergence is roughly achieved after the $\sim$ 40th samples.
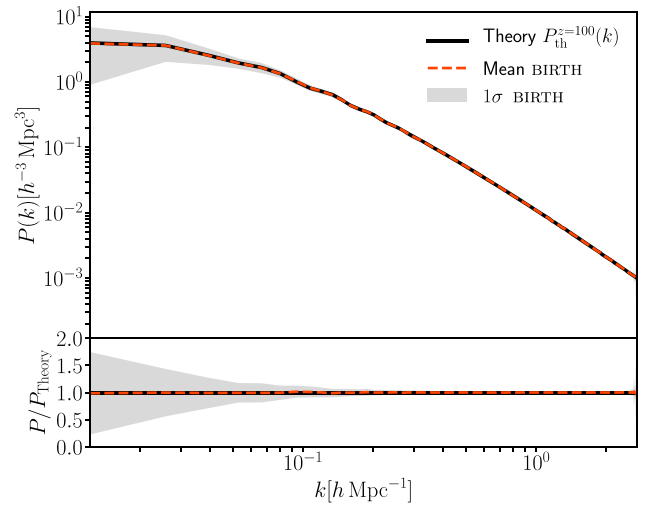

Figure 8. Mean power spectrum of the inferred initial density field $\delta(q)$ shown for the sub-volume ranging from $1.7 \leq z \leq 2.2$. The red-dashed line represents the mean power spectrum averaged over 6000 individual HMC realisations with 1σ standard deviation shown as a grey band. In the bottom panel, we show the ratio of the mean power spectrum and its uncertainty band with the theoretical ΛCDM prediction.

bias values allowed within 2σ confidence levels by the observations (see discussion in Section 4.2). The convergence behaviour of the power spectra from the Lagrangian density fields (for the best large-scale bias model), starting from a perfectly homogeneous density field $\delta(q) = 0$ is shown in Fig. 7. The reconstructions show unbiased power spectra already after the 40th iteration step. Up to this iteration, we consider the chain to be in the burn-in phase and not representing the target distribution of which we aim at drawing samples from. Furthermore, the top panel of Fig. 8 shows the mean power spectrum of the initial density field $\delta(q)$ at $z = 100$ averaged over 6000 samples, illustrating the 1σ standard deviation as a grey band. The bottom panel shows the corresponding ratios with the theoretical linear ΛCDM prediction. The reconstructed density fields show unbiased
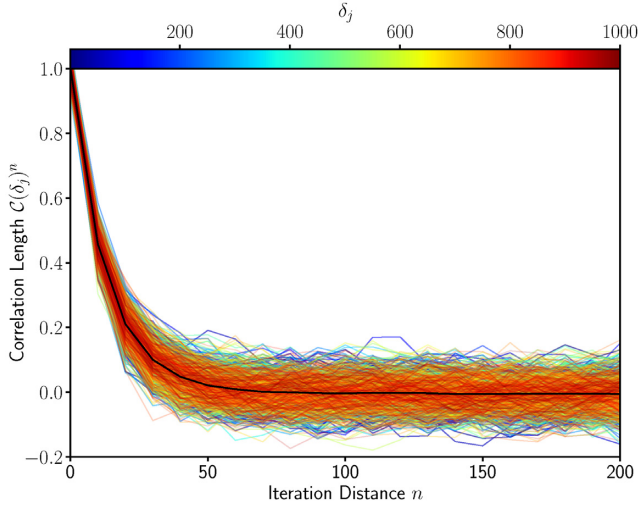
Figure 9. Correlation length $\mathcal{C}(\delta_j)^n$ of 1000 randomly chosen density field voxels δ_j, where the completeness is $w_j > 0$. We compute the correlation length over a sample of $N = 6500$ HMC realization with an iteration length $n \in [0...200]$ shown for the sub-volume ranging from $1.7 \leq z \leq 2.2$. The colour code indicates the correlation length for a particular δ_j while the black solid line represents the average over all density voxels. We consider $\mathcal{C}(\delta_j)^n < 0.1$ to be uncorrelated, which our Gibbs-sampling chain drops below after $n = 40$ iterations. Thus, each 40th realization is an independent sample.

power spectra over all scales, confirming an accurate bias treatment. The effect of a mishandled bias in our study is shown in Appendix A. We note that sampling the large-scale bias from sparse data sets as considered here with such a low data volume fraction is going to be inaccurate, although the Bayesian framework allows for it (see Granett et al. 2015; Jasche & Lavaux 2017).

Once the cosmology has been chosen, and the data input defined, there is only one free parameter in the COSMIC BIRTH code, which is the large-scale bias of the different populations. As shown in Section 4, there is a confidence region for the large-scale bias as a function of redshift for the galaxies considered in this study. Given the the low volume filling fraction of the data region with respect to the entire volume, which needs to be considered to keep the mode-coupling effects from large-scale modes low (see e.g. Sato et al. 2009; Takada & Hu 2013), the question arises whether the reconstructions are sensitive to bias. To validate the independence of the HMC samples, we calculate the correlation length for the inferred density field. The correlation length $\mathcal{C}(\delta_j)^n$ for a particular density voxel δ_j at an iteration distance n over N samples is given by

$$\mathcal{C}(\delta_j)^n = \frac{1}{N-n} \sum_{i=0}^{N-n} \frac{(\delta_j^i - \langle \delta_j \rangle)}{\sqrt{\sigma^2(\delta_j)}} \frac{(\delta_j^{i+n} - \langle \delta_j \rangle)}{\sqrt{\sigma^2(\delta_j)}}, \tag{23}$$

where $\langle \delta_j \rangle = \frac{1}{N} \sum_i \delta_j^i$ is the mean of the density voxel δ_j over N samples and $\sigma^2(\delta_j) = \frac{1}{N} \sum_i \left(\delta_j^i - \langle \delta_j \rangle \right)^2$ the corresponding variance. We show the correlation length in Fig. 9 for 1000 randomly chosen density voxels δ_j, with $j \in [1...1000]$ in the data region of our reconstructed volume. This demonstrates that we draw independent samples each $\sim$40th Gibbs-sampling iteration.

To quantify the Lagrangian to Eulerian mapping using ALPT within COSMIC BIRTH, we show the density distribution of the initial and final density fields in Fig. 10. On the same scale, on the x-axis the density field value δ is plotted against the normalized probability distribution. We can see that the density field in Lagrangian space follows closely a Gaussian distribution with mean
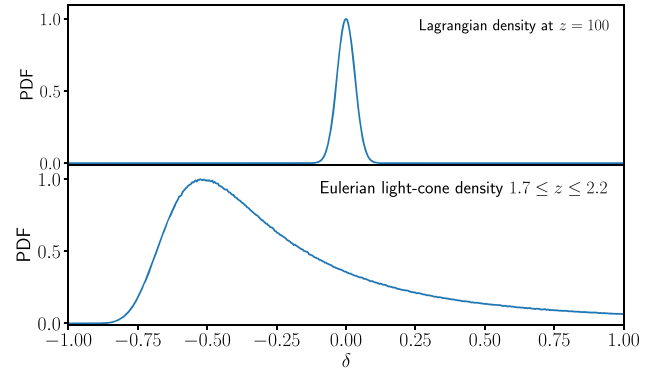

Figure 10. Histogram of the inferred density fields normalized to unity for $-1 \leq \delta \leq 1$. *The top panel* shows a histogram of the density at initial conditions (Lagrangian coordinates) at $z = 100$. *The bottom panel* shows a histogram of the light-cone density field in Eulerian coordinates in the range of $1.7 \leq z \leq 2.2$. The Lagrangian density on the top panel shows a closely Gaussian distribution with a negligible skewness of $s = 0.001$, while the Eulerian density on the bottom panel presents a highly skewed distribution with $s = 5.373$.

peaked at zero, and very small variance and skewness, $\mu \approx 0$, $\sigma^2 = 0.001$, and $s = 0.001$. The histogram for the Eulerian density field also has a mean value close to zero, $\mu \approx 0$, however, shows a variance of $\sigma^2 = 0.95$, and a skewness of $s = 5.373$. This is induced by gravity over cosmic time-scales and is in excellent agreement to previous findings studied in detail in Neyrinck (2013), comparing the displacement fields of second-order Lagrangian perturbation theory (2LPT) and N-body simulations at different redshifts. Especially at redshifts $z \geq 1$ the 2LPT displacements and the N-body results are in very good agreement and thus represent a very reasonable choice for this study (even more so, since we use ALPT).

5.3 Density inference results

The resulting reconstructions corresponding to our best large-scale bias evolution estimation are shown in Fig. 11 as slice plots. On the x-axis, we show the line-of-sight distance[4] as redshift z, and on the y-axis we show the declination (Dec.). The slices are shown with a thickness of $6 \, h^{-1}$ Mpc (averaging over three neighbouring cells). The corresponding convergence behaviour, power spectra, and matter statistics was discussed in Section 5.2 and shown in Figs 7–10.

We find homogeneously distributed initial cosmic density fields (see first panel of Fig. 11 for individual reconstructions of $1 + \boldsymbol{\delta}(\boldsymbol{q})$). This is further demonstrated in the statistical matter distribution shown in the upper panel of Fig. 10) with negligible skewness and kurtosis values, as expected for a Gaussian density field. Furthermore, the top panel of Fig. 11 shows no redshift evolution. This is in contrast with the second upper panel, in which the Eulerian density for the corresponding reconstruction $1 + \boldsymbol{\delta}(\boldsymbol{s})$ is shown. In fact, the corresponding matter statistics shown in the lower panel of Fig. 10 shows a highly non-Gaussian distribution. Further inspection of the second panel of Fig. 11, shows an increase of the matter fluctuations towards low redshifts, particularly enhanced through the appearance of large cosmic voids, depicted in dark blue. The corresponding

---

[4]We note that the line-of-sight distance corresponds to the redshift of the observations only for the light-cones. For the initial density fields $\boldsymbol{\delta}(\boldsymbol{q})$ at $z = 100$, x-axis only shows a distance measure.
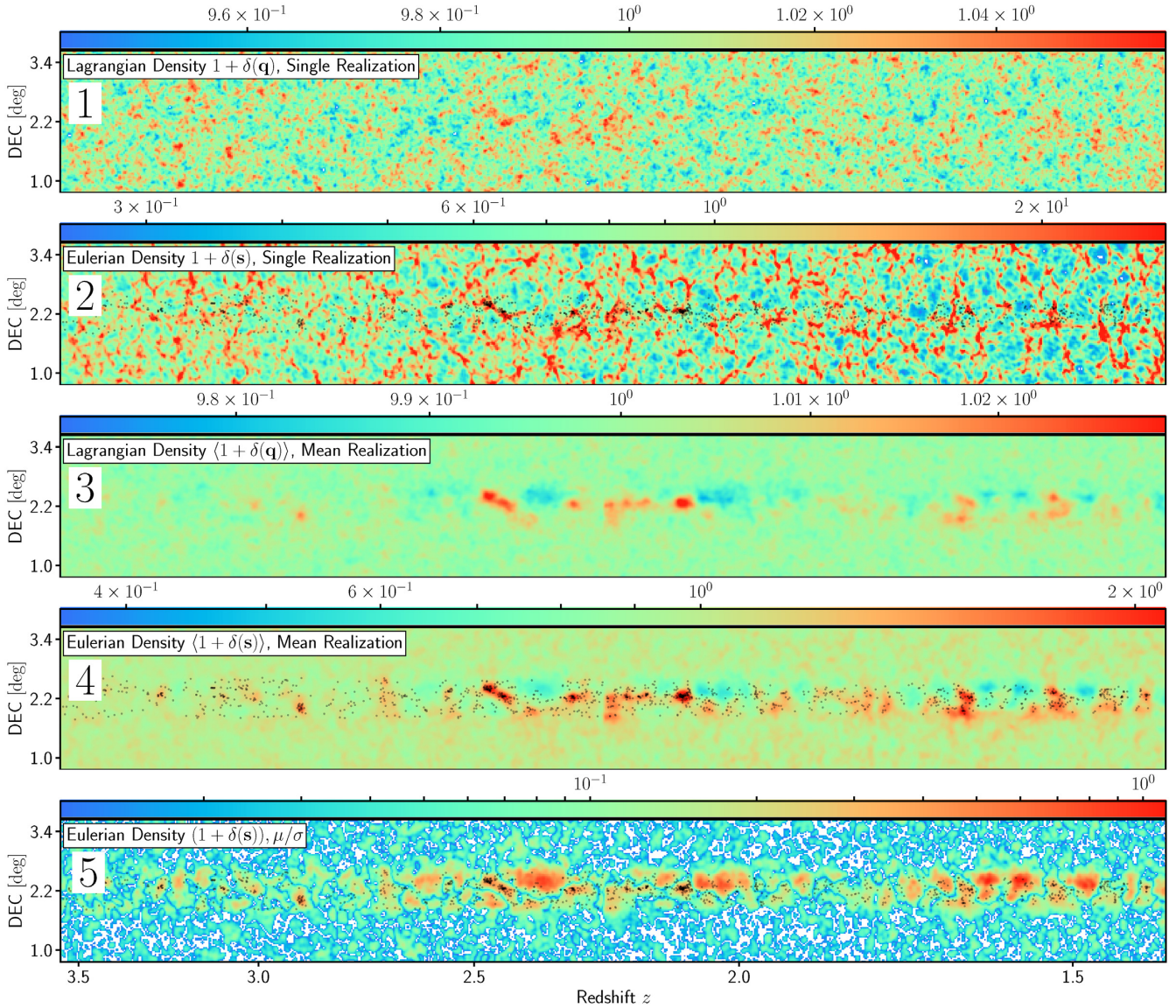
Figure 11. COSMIC BIRTH reconstructed density field $1 + \delta$ shown as slices in line-of-sight and declination coordinates with thickness of $6\,h^{-1}$ Mpc. Phase-space mapping with the corresponding tetra-hedra tesselation has been performed as mentioned in Section 2. The whole declination angle on the y-axis corresponds to $200\,h^{-1}$ Mpc in comoving coordinates. From top to bottom, the two upper slice plots, panels 1 and 2, show first a single realization the Lagrangian initial density field $1 + \delta(\mathbf{q})$ at $z = 100$ and secondly a light-cone realization at Eulerian redshift-space $1 + \delta(\mathbf{s})$ with the individual galaxy positions plotted on top. Panels 3 and 4 represent the mean distributions of the inferred initial Lagrangian $\langle 1 + \delta(\mathbf{q}) \rangle$ at $z = 100$ and final Eulerian $\langle 1 + \delta(\mathbf{s}) \rangle$ density fields averaged over 6000 HMC realizations, respectively, where on top of the Eulerian density field the galaxy positions are plotted with the black dots. Finally, we show in panel 5 the signal-to-noise ratio μ/σ of the Eulerian density field with the galaxy positions plotted on top.

ensemble averages (over 6000 Gibbs-sampling iterations) to panels 1 and 2 are shown in the panels 3 and 4, $\langle 1 + \delta(\mathbf{q}) \rangle$ and $\langle 1 + \delta(\mathbf{s}) \rangle$, respectively. In particular, the average over the ensemble of Eulerian density fields (i.e. the expected dark matter field from a Bayesian calculation) shows a high correlation with the galaxy field, as it should happen, when the initial cosmic density field is accurately re-covered. One can also appreciate the vanishing fluctuations in regions with low completeness, in concordance with a Bayesian analysis (see for a comparison, as an example the completeness of VUDS depicted in Fig. 5). Panel 5 shows the signal-to-noise ratio obtained through the ratio of the mean density over its standard deviation μ/σ. As expected, we find a higher signal-to-noise ratio in the data region. Interestingly, cosmic voids are particularly prominent in this

measure. The missing data or equivalently, empty window function regions increase the uncertainty of the corresponding density cells and thus the variance of the density field (see equation 6). This would be reflected in the signal-to-noise ratio, shown in the bottom panel of Fig. 11. We however see no significant decrease in the signal-to-noise ratio at the gaps of the VUDS survey.

To further assess the matter distribution in the reconstructions, we show tomographic slice plots for various RA ranges (the grey vertical stripes) in Fig. 12, showing the survey footprints on the left-hand side and the corresponding density reconstructions next to it on the right. This permits us to evaluate the extension of the proto-clusters and cosmic voids. A comparison between the Lagrangian and Eulerian ensemble average reconstructions (middle
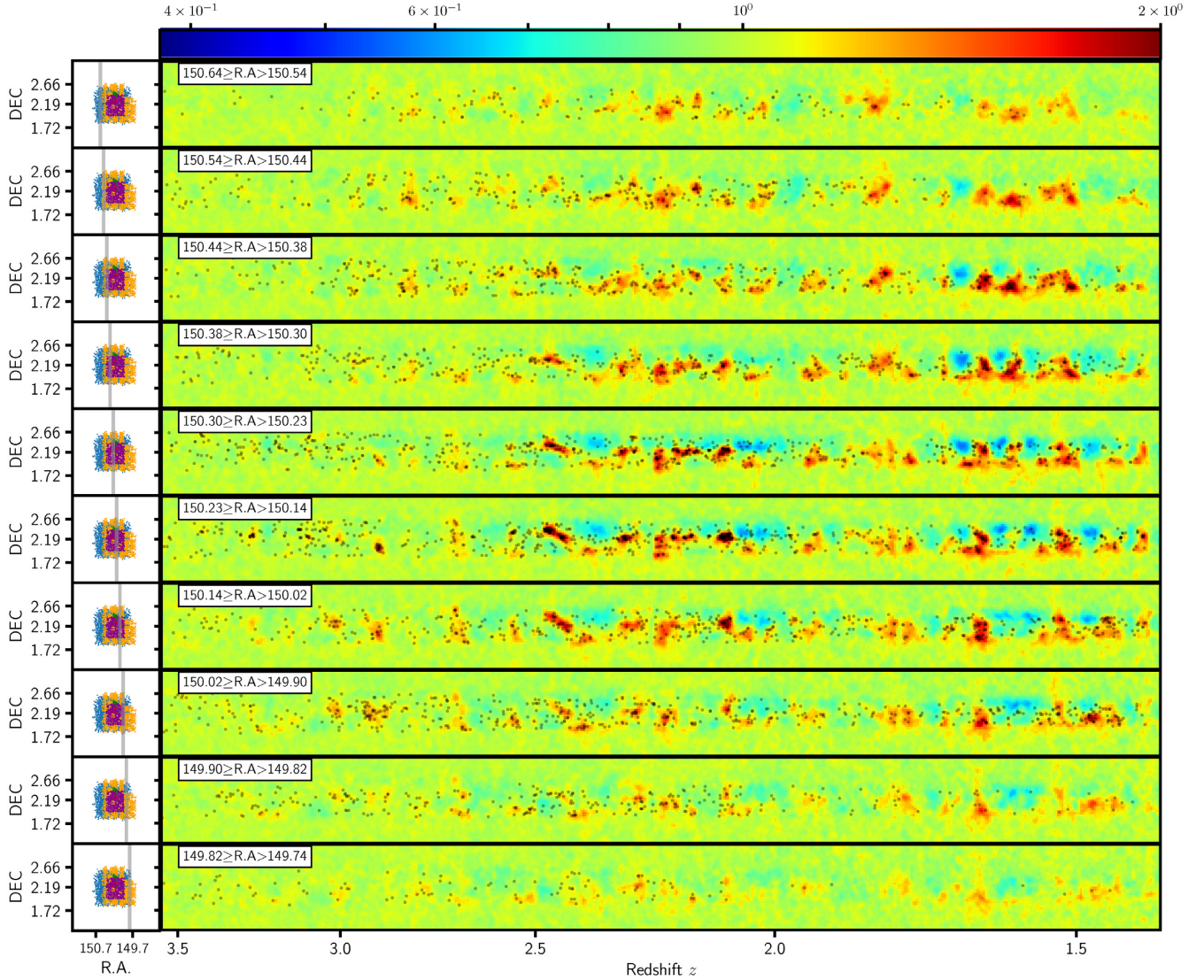
Figure 12. Tomographic slices of the reconstructed matter density field in the declination-redshift plane. On the left, we show the angular footprint of the five surveys in the same colour code as Fig. 2 with a grey line showing the RA thickness of the corresponding slice. On the right, the mean light-cone density reconstructions are shown.

panels in Fig. 11) shows higher density fluctuations in Eulerian space, and small displacements of the center of mass of the proto-clusters towards high redshifts. Focusing on the lowest redshift regions one can observe stronger variations of the shape of the proto-clusters.

To further investigate this, we zoom into a prominent proto-cluster region with $2.28 \leq z \leq 2.51$ and $1.6° \leq$ DEC. $\leq 2.9°$ (see Fig. 13). This comparison shows three high-density proto-clusters at $z \sim 2.45$, $z \sim 2.4$, and $z \sim 2.3$ growing throughout the cosmological timescale starting from $z = 100$ down to their light-cone redshifts. At first glance, the proto-clusters seem to grow in place, with negligible displacements, in accordance with linear perturbation theory, describing the growth of perturbation fixed in comoving frame between the initial redshift z_q to the final redshift z_f with $\delta(\mathbf{r}, z_f) = \delta(\mathbf{r}, z_q) \frac{D(z_f)}{D(z_q)}$, where $D(z)$ is the linear growth factor. Linear theory from $z_q = 100$ to the redshift of our observed light-cone ($\sim z_f = 2.3$) predicts a growth by a factor of approximately 30, which is consistent with the colour bar of the first and second panel of Fig. 11. However, a more detailed inspection reveals a complex

non-spherical accumulation of mass when comparing the middle and the right-hand panels. This can be well appreciated when comparing to the Gaussian smoothed galaxy field shown on the left-hand panel. We also find from this comparison how the two proto-clusters on the left, which appear as separate entities in the left-hand panel, are actually connected, most likely having a dark matter bridge in between their respective galaxy distributions. It is also interesting to see the accumulation of matter through the action of gravity from a ring-like shape cloud to a spherical overdensity region in the region centred at $z \sim 2.38$ and Dec. $\sim 1.9°$, comparing the Lagrangian and Eulerian density fields. We also focus on the prominent $z \sim 2.1$ proto-cluster and show how this region is reconstructed within different individual realizations in Appendix B. This Bayesian analysis taking the completeness into account, also reveals regions that are more likely to be true cosmic voids. In particular, the right-hand panel shows a deep void towards higher declination angles in the redshift range $2.3 < z < 2.4$, which feeds the overdensity peaks on its left and right. This is in agreement with the cosmic void initially discovered from the three dimensional Ly α forest tomography (Krolewski et al.
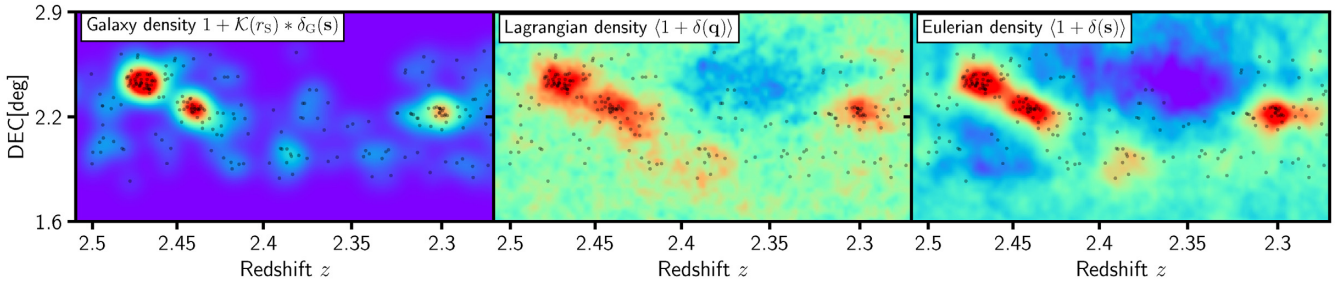
Figure 13. Density plots for a zoomed region. On the left-hand panel, the galaxy density $1 + \mathcal{K}(r_S) * \delta_G(s)$ is shown convoluted with a Gaussian smoothing kernel $\mathcal{K}(r_S)$ with smoothing radius of $r_S = 2\,h^{-1}$ Mpc, followed by the mean density field in initial (middle panel) $1 + \delta(q)$ and final (right-hand panel) conditions $1 + \delta(s)$ for the same slice as shown in Fig. 11. Overplotted with the black dots are shown the corresponding galaxy positions.

2018). Another void region is forming directly under the proto-clusters at $z \sim 2.45$ and $z \sim 2.4$. We will provide a more detailed analysis of the various structures in the reconstructed density field in a subsequent paper.

6 CONCLUSIONS AND DISCUSSION

In this work, we presented for the first time a comprehensive multisurvey reconstruction effort of the primordial and evolved density fields with the `COSMIC BIRTH` algorithm performed in the COSMOS field during the epoch of Cosmic Noon ($1.4 \leq z \leq 3.6$). To our knowledge, this is the ever attempted large-scale structure analysis of its kind at high redshift, probing the quasi-linear regime of gravitational structure formation, before non-linear shell-crossing started dominating the emerging cosmic web.

We combined the data from five spectroscopic galaxy surveys in the COSMOS field, which have partially overlapping footprints, and do not share a common observing strategy. Therefore, we had to make a special effort in estimating the angular selection function for each survey, based on the the individual targeting strategies for the parent photometric catalogues, which was missing for these surveys.

Also, we applied for the first time a multitracer and multisurvey likelihood formalism in Lagrangian space within a Bayesian inference framework. This allowed us to combine surveys with different selection functions, galaxy bias, number densities, and redshift ranges, connected through the underlying dark matter distribution. Although earlier works have presented reconstructions using multiple galaxy populations, they, however, shared a unique survey geometry. Therefore, our new method is more general and has a wider range of possible applications. Since we expect a correlation between the spatial distribution of the galaxy catalogues, a covariance term would in principle be necessary when different observations are combined within one joint likelihood analysis. We bypass this problem by first mapping all tracers to Lagrangian coordinates. This allows us to assume Poisson likelihoods, since only in the homogeneous epoch before gravity coupled separated spatial regions, identical and independently distributed large-scale structure tracers can be assumed.

Despite of the large number of surveys, the `COSMIC BIRTH` code showed an efficient performance, converging within about 40 iterations, and showing low correlation lengths of about 40 Gibbs-sampling iterations.

The resulting reconstructions reveal for the first time a holistic view on the matter density field and its primordial fluctuations jointly inferred from five spectroscopic surveys.

We also revised the bias of star-forming galaxies towards high redshifts, finding some moderate evidence for a stronger evolution than the one described by passive evolution.

The inferred density fields have a large number of potential applications. In particular, we have found several high-density regions across the whole reconstructed volume. We successfully reconstructed a number of observationally known proto-cluster regions previously reported by Cucciati et al. (2014), Chiang et al. (2015), Diener et al. (2015), Casey et al. (2015), Lee et al. (2016), Wang et al. (2016), Nanayakkara et al. (2016), and Darvish et al. (2020). However, these previous proto-cluster studies were typically performed using individual galaxy surveys directly on the observed galaxy positions in redshift-space. Major improvements have been done combining two surveys and statistically sampling the redshifts in 2D slices including a Voronoi tessellation to estimate the galaxy density field (Cucciati et al. 2018). Still, the mass estimates are prone to projection effects due to their peculiar velocities (Kaiser 1984, 1987), and also assuming a velocity dispersion measure which is not valid for non-virialized objects at high redshifts. In a subsequent paper, we will carry out a more detailed analysis of these structures. Since we have reconstructed the initial conditions of the COSMOS volume, we will be able to run constrained *N*-body simulations based on the inferred initial conditions, enabling us to study the full cosmic evolution of the proto-clusters in detail (Ata et al., in preparation). This will, for example, permit us to direct model the late-time properties of the galaxy clusters that will coalesce from the proto-clusters observed in the COSMOS surveys. Furthermore, the Bayesian formalism will permit us to directly quantify the uncertainty of late-time cluster properties.

In addition to studying the large-scale structure evolution, there are many potential applications for the contemporary density field derived in COSMOS. Our density map could be used to directly address the question of galaxy evolution in the context of environment (Nuza et al. 2014), without having to use contrived statistics (e.g. counts in cylinders, *N*-nearest neighbours) to define the galaxy environment (Cooper et al. 2008; Koyama et al. 2013; Kawinwanichakij et al. 2017; Ji et al. 2018; Muldrew et al. 2018). More remarkably, by treating the observed COSMOS galaxies as tracer particles in the constrained *N*-body simulations based on these reconstruction, we will also be able to track them to their eventual $z = 0$ environments and thus link them directly with well-studied trends in the Local Universe (e.g. Kauffmann et al. 2004).

Another clear application is to compare the dark matter reconstructions from galaxy tracers with reconstructions obtained from hydrogen Ly α forest tomography observations in the CLAMATO survey (Lee et al. in preparation). This will allow to directly test the fluctuating Gunn–Peterson approximation that posits a monotonic

relationship between Lyα forest absorption and the underlying density field. While the CLAMATO data are reconstructed moderate-resolution, low-S/N LBG spectra, there also exist several high-resolution absorption spectra that have been observed of bright quasars in the field. On these high-resolution spectra, we will aim also to carry out an analysis of the line widths and column densities as a function of the underlying matter density, which will shed light into a recent study which suggested that the thermal properties of the IGM varies between low-density and high-density regions (Rorai et al. 2018).

The method applied in this study is also interesting for the Galaxy Evolution component of the planned Subaru PFS Subaru Strategy Program, which will target $\sim 12°-15°$ of deep spectroscopy over three continuous fields, i.e. an order-of-magnitude larger area than COSMOS. The galaxy number densities are also very comparable, or better, than those reconstructed in this paper. For the low-redshift NIR-selected sample at $z \sim 1-1.5$, the PFS will obtain spectra for galaxies at a number density of $n \approx 3 \times 10^{-3}\,h^3\,\mathrm{Mpc}^{-3}$, slightly better than the FMOS–COSMOS sample that has $n \sim 2 \times 10^{-3}\,h^3\,\mathrm{Mpc}^{-3}$ at $z \sim 1.7$; This should enable direct cosmic web analysis from the galaxy reconstructions alone. Closer to the peak of Cosmic Noon at $z \sim 2-3$, an optically selected LBG sample is planned to yield number densities equivalent to VUDS or zCOSMOS-deep ($n \sim 3 \times 10^{-4}\,h^{-3}\,\mathrm{Mpc}^{-3}$). While this is less than the combined survey sample used in this paper, it should still be sufficient to identify and characterise proto-clusters within the volume.

As wide-field spectroscopic surveys probe ever deeper into cosmic history, the utility of density reconstruction techniques will become increasingly important to fully exploit the rich scientific possibilities that will open up.

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Abel T., Hahn O., Kaehler R., 2012, MNRAS, 427, 61
Abidi M. M., Baldauf T., 2018, J. Cosmol. Astropart. Phys., 2018, 029
Alam S. et al., 2017, MNRAS, 470, 2617
Allen S. W., Evrard A. E., Mantz A. B., 2011, ARA&A, 49, 409
Amara A. et al., 2012, MNRAS, 424, 553
Amendola L. et al., 2018, Living Rev. Relativ., 21
Anderson L. et al., 2012, MNRAS, 427, 3435
Ata M., Kitaura F.-S., Müller V., 2015, MNRAS, 446, 4250
Ata M. et al., 2017, MNRAS, 467, 3993
Ata M. et al., 2018, MNRAS, 473, 4773
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Beckwith S. V. W. et al., 2006, AJ, 132, 1729
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Belfiore F. et al., 2017, MNRAS, 466, 2570
Bernardeau F., 1994, ApJ, 427, 51
Bode P., Ostriker J. P., Turok N., 2001, ApJ, 556, 93
Bond J. R., Kofman L., Pogosyan D., 1996, Nature, 380, 603
Bouchet F. R., Juszkiewicz R., Colombi S., Pellat R., 1992, ApJ, 394, L5
Bouchet F. R., Colombi S., Hivon E., Juszkiewicz R., 1995, A&A, 296, 575
Brammer G. B. et al., 2012, ApJS, 200, 13
Capak P. et al., 2007, ApJS, 172, 99
Casey C. M. et al., 2015, ApJ, 808, L33
Casey C. M., Hodge J., Zavala J. A., Spilker J., da Cunha E., Staguhn J., Finkelstein S. L., Drew P., 2018, ApJ, 862, 78
Catelan P., 1995, MNRAS, 276, 115
Chiang Y. K., Overzier R., Gebhardt K., 2013, ApJ, 779, 127
Chiang Y. K. et al., 2015, ApJ, 808, 37
Chiang Y.-K., Overzier R. A., Gebhardt K., Henriques B., 2017, ApJ, 844, L23
Cirasuolo M. et al., 2014, in McLean I. S., Ramsay S. K., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham, p. 84460S
Cochrane R. K., Best P. N., Sobral D., Smail I., Wake D. A., Stott J. P., Geach J. E., 2017, MNRAS, 469, 2913
Cohn J. D., White M., 2008, MNRAS, 385, 2025
Coles P., Jones B., 1991, MNRAS, 248, 1
Cooper M. C., Tremonti C. A., Newman J. A., Zabludoff A. I., 2008, MNRAS, 390, 245
Costanzi Alunno Cerbolini M., Sartoris B., Xia J.-Q., Biviano A., Borgani S., Viel M., 2013, J. Cosmol. Astropart. Phys., 2013, 20
Cucciati O. et al., 2014, A&A, 570, A16
Cucciati O. et al., 2018, A&A, 619, A49
Daddi E., Cimatti A., Renzini A., Fontana A., Mignoli M., Pozzetti L., Tozzi P., Zamorani G., 2004, ApJ, 617, 746
Darvish B., Mobasher B., Sobral D., Scoville N., Aragon-Calvo M., 2015, ApJ, 805, 121
Darvish B. et al., 2020, ApJ, 892, 8
Delubac T. et al., 2017, MNRAS, 465, 1831
de Jong R. S. et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham, p. 84460T
Diener C. et al., 2013, ApJ, 765, 109
Diener C. et al., 2015, ApJ, 802, 31
Duane S., Kennedy A. D., Pendleton B. J., Roweth D., 1987, Phys. Lett. B, 195, 216
Durkalec A. et al., 2015, A&A, 583, A128
Durkalec A. et al., 2018, A&A, 612, A42
Eisenstein D. J. et al., 2005, ApJ, 633, 560
Eisenstein D. J., Seo H.-J., Sirko E., Spergel D. N., 2007, ApJ, 664, 675
Favole G. et al., 2016, MNRAS, 461, 3421
Fry J. N., 1996, ApJ, 461, L65
Fry J. N., Peebles P. J. E., 1978, ApJ, 221, 19
Gao L., White S. D. M., Jenkins A., Frenk C. S., Springel V., 2005, 363, 379
Granett B. R. et al., 2015, A&A, 583, A61
Guo H. et al., 2019, ApJ, 871, 147
Guzzo L. et al., 2014, A&A, 566, A108
Hada R., Eisenstein D. J., 2018, MNRAS, 478, 1866
Hahn O., Abel T., Kaehler R., 2013, MNRAS, 434, 1171
Hasinger G. et al., 2018, ApJ, 858, 77
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663
Hernández-Sánchez M., Kitaura F.-S., Ata M., Vecchia C. D., 2019, preprint (arXiv:1911.02667)
Heß S., Kitaura F.-S., Gottlöber S., 2013, MNRAS, 435, 2065
Horowitz B., Lee K.-G., White M., Krolewski A., Ata M., 2019, ApJ, 887, 61
Ilbert O. et al., 2005, A&A, 439, 863
Ilbert O. et al., 2013, A&A, 556, A55
Jasche J., Kitaura F. S., 2010, MNRAS, 407, 29
Jasche J., Lavaux G., 2017, A&A, 606, A37

Jasche J., Lavaux G., 2019, A&A, 625, A64

Jasche J., Wandelt B. D., 2013, MNRAS, 432, 894

Jasche J., Leclercq F., Wandelt B. D., 2015, J. Cosmol. Astropart. Phys., 2015, 036

Ji Z., Giavalisco M., Williams C. C., Faber S. M., Ferguson H. C., Guo Y., Liu T., Lee B., 2018, ApJ, 862, 135

Kaasinen M., Bian F., Groves B., Kewley L. J., Gupta A., 2017, MNRAS, 465, 3220

Kaiser N., 1984, ApJ, 284, L9

Kaiser N., 1987, MNRAS, 227, 1

Kashino D. et al., 2017, ApJ, 843, 138

Kashino D. et al., 2019, ApJS, 241, 10

Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, MNRAS, 303, 188

Kauffmann G., White S. D. M., Heckman T. M., Ménard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004, MNRAS, 353, 713

Kawinwanichakij L. et al., 2017, ApJ, 847, 134

Kennicutt R. C., Jr, 1998, ARA&A, 36, 189

Kimura M. et al., 2010, PASJ, 62, 1135

Kitaura F.-S., 2013, MNRAS, 429, L84

Kitaura F.-S., Angulo R. E., 2012, MNRAS, 425, 2443

Kitaura F.-S., Enßlin T. A., 2008, MNRAS, 389, 497

Kitaura F.-S., Hess S., 2013, MNRAS, 435, L78

Kitaura F.-S., Jasche J., Li C., Enßlin T. A., Metcalf R. B., Wandelt B. D., Lemson G., White S. D. M., 2009, MNRAS, 400, 183

Kitaura F.-S., Jasche J., Metcalf R. B., 2010, MNRAS, 403, 589

Kitaura F.-S., Gallerani S., Ferrara A., 2012a, MNRAS, 420, 61

Kitaura F.-S., Angulo R. E., Hoffman Y., Gottlöber S., 2012b, MNRAS, 425, 2422

Kitaura F.-S., Erdoğdu P., Nuza S. E., Khalatyan A., Angulo R. E., Hoffman Y., Gottlöber S., 2012c, MNRAS, 427, L35

Kitaura F.-S., Yepes G., Prada F., 2014, MNRAS, 439, L21

Kitaura F.-S., Gil-Marín H., Scóccola C. G., Chuang C.-H., Müller V., Yepes G., Prada F., 2015, MNRAS, 450, 1836

Kitaura F.-S., Ata M., Angulo R. E., Chuang C.-H., Rodríguez-Torres S., Monteagudo C. H., Prada F., Yepes G., 2016, MNRAS, 457, L113

Kitaura F.-S., Ata M., Rodriguez-Torres S. A., Hernandez-Sanchez M., Balaguera-Antolinez A., Yepes G., 2019, preprint (arXiv:1911.00284)

Kollmeier J. A., Weinberg D. H., Davé R., Katz N., 2003, ApJ, 594, 75

Kovač K. et al., 2010, ApJ, 708, 505

Koyama Y. et al., 2013, MNRAS, 434, 423

Kravtsov A. V., Borgani S., 2012, ARA&A, 50, 353

Kriek M. et al., 2015, ApJS, 218, 15

Krolewski A. et al., 2018, ApJ, 861, 60

Laigle C. et al., 2016, ApJS, 224, 24

Lavaux G., Jasche J., Leclercq F., 2019, preprint (arXiv:1909.06396)

Le Fèvre O. et al., 2015, A&A, 576, A79

Lee K.-G., Hennawi J. F., White M., Croft R. A., Ozbek M., 2014a, ApJ, 788, 49

Lee K.-G. et al., 2014b, ApJ, 795, L12

Lee K.-G. et al., 2016, ApJ, 817, 160

Lee K.-G. et al., 2018, ApJS, 237, 31

LeFevre O. et al., 2003, in Iye M., Moorwood A. F. M., eds, Proc. SPIE Conf. Ser. Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-Based Telescopes. SPIE, Bellingham, p. 1670

Lemaux B. C. et al., 2012, ApJ, 745, 106

Lemaux B. C. et al., 2018, A&A, 615, A77

Lemaux B. C. et al., 2019, MNRAS, 490, 1231

Levi M. et al., 2013, preprint (arXiv:1308.0847)

Lilly S. J. et al., 2007, ApJS, 172, 70

Lilly S. J. et al., 2009, ApJS, 184, 218

Madau P., Dickinson M., 2014, ARA&A, 52, 415

McLean I. S. et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham, p. 84460J

Modi C., Castorina E., Seljak U., 2017, MNRAS, 472, 3959

Mohayaee R., Mathis H., Colombi S., Silk J., 2006, MNRAS, 365, 939

Monaco P., Efstathiou G., 1999, MNRAS, 308, 763

Muldrew S. I., Hatch N. A., Cooke E. A., 2018, MNRAS, 473, 2335

Nanayakkara T. et al., 2016, ApJ, 828, 21

Neal R. M., 1993, in Technical Report CRG-TR-93-1 No. September. University of Toronto

Neal R. M., 2011, in Brooks S., Gelman A., Jones G. L., Meng X.-L., eds, Handbook of Markov Chain Monte Carlo. Chapman & Hall / CRC Press, London, p. 113

Newman A. B. et al., 2020, ApJ, 891, 147

Neyrinck M. C., 2013, MNRAS, 428, 141

Neyrinck M. C., 2016, MNRAS, 455, L11

Neyrinck M. C., Szapudi I., Szalay A. S., 2009, ApJ, 698, L90

Neyrinck M. C., Aragón-Calvo M. A., Jeong D., Wang X., 2014, MNRAS, 441, 646

Nusser A., Davis M., 1994, ApJ, 421, L1

Nuza S. E., Kitaura F.-S., Heß S., Libeskind N. I., Müller V., 2014, MNRAS, 445, 988

Okada H. et al., 2016, PASJ, 68, 47

Patrick Bos E. G., Kitaura F. S., van de Weygaert R., 2019, MNRAS, 488, 2573

Peebles P. J. E. P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ

Percival W. J., Cole S., Eisenstein D. J., Nichol R. C., Peacock J. A., Pope A. C., Szalay A. S., 2007, MNRAS, 381, 1053

Pezzotta A. et al., 2017, A&A, 604, A33

Pichon C., Vergely J. L., Rollinde E., Colombi S., Petitjean P., 2001, MNRAS, 326, 597

Porqueres N., Jasche J., Lavaux G., Enßlin T., 2019, A&A, 630, A151

Reid B. A., White M., 2011, MNRAS, 417, 1913

Rorai A., Carswell R. F., Haehnelt M. G., Becker G. D., Bolton J. S., Murphy M. T., 2018, MNRAS, 474, 2871

Sato M., Hamana T., Takahashi R., Takada M., Yoshida N., Matsubara T., Sugiyama N., 2009, ApJ, 701, 945

Schmittfull M., Baldauf T., Zaldarriaga M., 2017, Phys. Rev. D, 96, 023505

Schmittfull M., Simonović M., Assassi V., Zaldarriaga M., 2019, Phys. Rev. D, 100, 043514

Scodeggio M. et al., 2018, A&A, 609, A84

Scoville N. et al., 2007, ApJS, 172, 1

Scoville N. et al., 2013, ApJS, 206, 3

Shi K. et al., 2019, ApJ, 879, 9

Silverman J. D. et al., 2015, ApJS, 220, 12

Skrutskie M. F. et al., 2006, AJ, 131, 1163

Smolčić V. et al., 2017, A&A, 597, A4

Somerville R. S., Davé R., 2015, ARA&A, 53, 51

Sorce J. G. et al., 2016, MNRAS, 455, 2078

Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelberger K. L., 1996, ApJ, 462, L17

Steidel C. C., Shapley A. E., Pettini M., Adelberger K. L., Erb D. K., Reddy N. A., Hunt M. P., 2004, ApJ, 604, 534

Straatman C. M. S. et al., 2016, ApJ, 830, 51

Takada M., Hu W., 2013, Phys. Rev. D, 87, 123504

Takada M. et al., 2014, PASJ, 66, R1

Tegmark M., Peebles P. J. E., 1998, ApJ, 500, L79

Thomas S. A., Abdalla F. B., Lahav O., 2011, Phys. Rev. Lett., 106, 241301

Tinker J. L., 2007, MNRAS, 374, 477

Tomczak A. R. et al., 2019, MNRAS, 484, 4695

Tully R. B., Courtois H., Hoffman Y., Pomarède D., 2014, Nature, 513, 71

Wang H., Mo H. J., Jing Y. P., Guo Y., van den Bosch F. C., Yang X., 2009, MNRAS, 394, 398

Wang H., Mo H. J., Yang X., van den Bosch F. C., 2013, ApJ, 772, 63

Wang H., Mo H. J., Yang X., Jing Y. P., Lin W. P., 2014, ApJ, 794, 94

Wang T. et al., 2016, ApJ, 828, 56

Watson W. A., Iliev I. T., D'Aloisio A., Knebe A., Shapiro P. R., Yepes G., 2013, MNRAS, 433, 1230

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Yahil A., Strauss M. A., Davis M., Huchra J. P., 1991, ApJ, 372, 380

Yuan T. et al., 2014, ApJ, 795, L20

Zaroubi S., Hoffman Y., Fisher K. B., Lahav O., 1995, ApJ, 449, 446
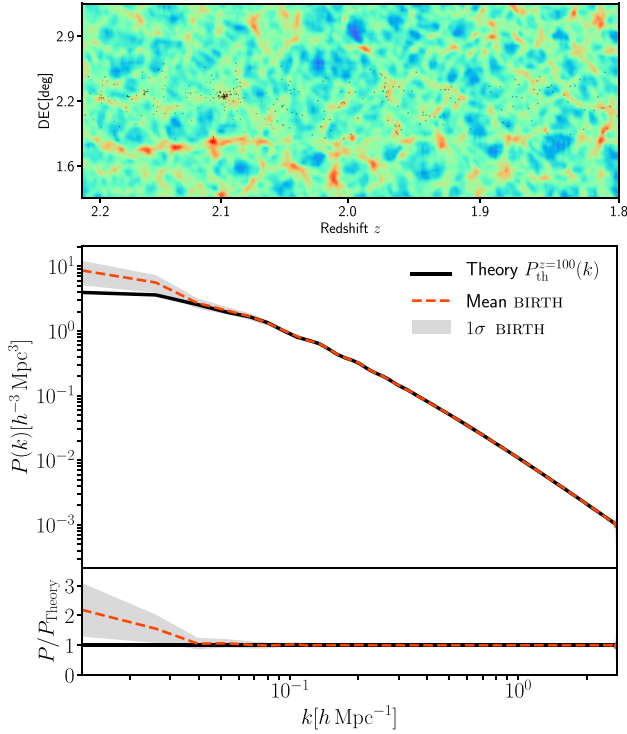
Zucca E. et al., 2009, A&A, 508, 1217

Figure A1. Results from COSMIC BIRTH runs with a bias exceeding the mean measurement given by the literature by 2σ in the redshift range $1.8 < z < 2.2$. *Top*: Slice plot of the light-cone density field $\delta(s)$ with the galaxy positions plotted on top represented by the black dots. *Bottom*: Power spectra of the corresponding initial density fields $\delta(q)$. We show the mean (the blue-dashed line) and the standard deviation (the grey band) for 1000 HMC realizations, including the ratio with the theoretical ΛCDM power spectrum in the panel below.

APPENDIX A: IMPACT OF AN INACCURATE BIAS TREATMENT

Even though the data region within the reconstructed mesh grid only occupies a small fraction, wrong galaxy bias estimates can

have a strong impact on the accuracy of the reconstructions, as we demonstrate here. In particular, we perform a COSMIC BIRTH run on a sub-volume covering the redshift range of $1.8 < z < 2.2$ with the data described in Section 3. We assume a large-scale bias ~ 60 per cent systematically higher than the one from our regular runs, as presented in Fig. 6 and discussed in Section 2.2.1. This bias exceeds the mean observational measurements in that redshift range by 2σ, and therefore represents a conservative upper bias limit (see the measurement by Kashino et al. 2017, extrapolated to higher redshifts with passive evolution in Fig. 6).

The results for this test are represented in Fig. A1. On the top, we show the inferred density field $\delta(s)$ with the galaxy positions plotted on top represented by the black dots. Although the reconstructed structures match the galaxy positions, a deeper inspection shows that the data region around DEC. $= 2.2$ is underweighted, due to the overestimated bias value and artificially high-density structures at the edge of the data region are formed. In the bottom plot, we show the mean power spectrum (the red-dashed) and the 1σ standard deviation (the grey band) of the inferred initial density fields for 1000 samples after convergence including the ratio of the inferred power spectra with the theoretical ΛCDM prediction below. Considering the low volume filling fraction, unbiased power spectra with the theoretical one are expected. However, a clear excess of power can be seen at the lowest modes, excluding the proposed bias values in this test. A thorough analysis of the bias, will be presented in a forthcoming paper.

APPENDIX B: DIFFERENT INDIVIDUAL CLUSTER RECONSTRUCTIONS

We focus on the prominent $z \sim 2.1$ (mainly seen in ZFIRE) proto-cluster and analyse the variance within the chain by looking at the result of 10 individual HMC reconstructions, which we choose randomly. We find on average a $\delta_{sum} = 841 \pm 121$ when we sum all cells in a $10\,h^{-1}$ Mpc cubic mask around the centre of the proto-cluster.
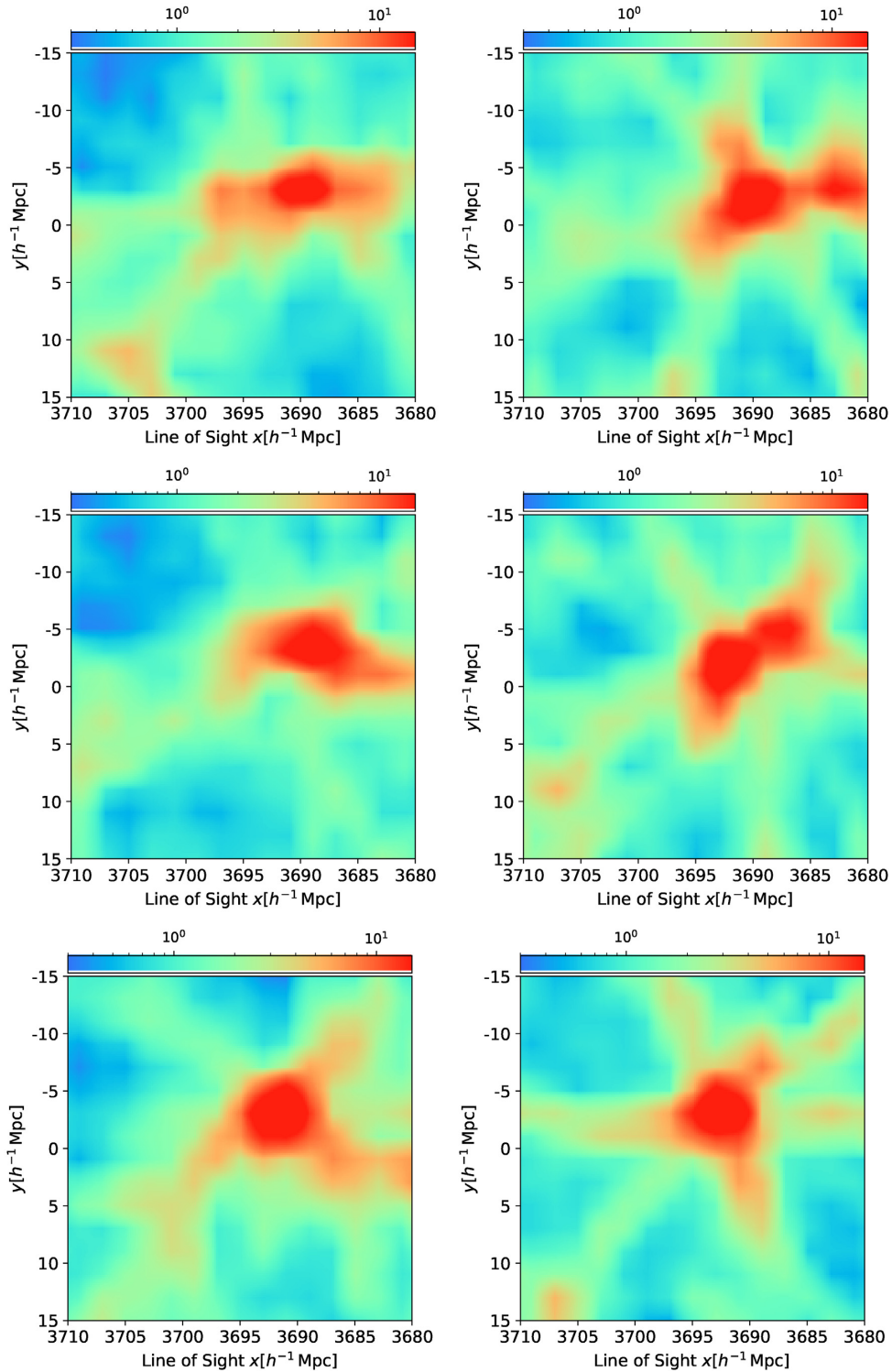
Figure B1. Cluster reconstruction among different, randomly chosen HMC realizations.

This paper has been typeset from a TEX/LATEX file prepared by the author.